



# Durham E-Theses

---

## *How moral knowledge motivates: a practical reason account*

PROCTOR, DUNCAN,EDWARD

### How to cite:

---

PROCTOR, DUNCAN,EDWARD (2010) *How moral knowledge motivates: a practical reason account*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/582/>

### Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in Durham E-Theses
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full Durham E-Theses policy](#) for further details.

## DUNCAN EDWARD PROCTOR

### How moral knowledge motivates: a practical reason account

When we make moral judgements and act morally we recognise and respond to reasons that are there whether we recognise them or not. This is the claim defended in this thesis. It has two aspects. The first is that acts of moral judgement aspire, sometimes successfully, to moral knowledge. This is moral cognitivism. The second is that moral truths report reasons for action. In responding appropriately to these reasons we are motivated to action. This is the practicality of morality. Hence, it is claimed, there is a moral reality that we respond to in both cognition and action. Adopting a practical reason approach, I argue that the objectivity and practicality of morality are not in conflict, but are linked by the idea of a practical reason. The moral truths that we can have knowledge of *are* the truths about the reasons for action that morality provides. I argue for this claim by showing why we should reject Humean ways of thinking about motivation and practical reason and embrace a broadly Kantian account. I argue that this account is compatible with seeing moral reasons as contributory rather than decisive or overriding. I also show how this account enables moral cognitivists to respond convincingly to arguments advanced by non-cognitivists.

# HOW MORAL KNOWLEDGE MOTIVATES

## A PRACTICAL REASON ACCOUNT

DUNCAN EDWARD PROCTOR

DOCTORAL THESIS

DEPARTMENT OF PHILOSOPHY

DURHAM UNIVERSITY

2010

## CONTENTS

---

<i>Abstract</i>	1
<i>Acknowledgements</i>	8
Chapter 1: Moral Cognitivism	9
Chapter 2: Moral Judgement Internalism	33
Chapter 3: Commonsense and the Humean Theory	61
Chapter 4: Platts' Dilemma	91
Chapter 5: Smith's Argument for the Humean Theory	115
Chapter 6: Normativity and the Humean Theory	143
Chapter 7: Normativity and Subjective Motivations	169
Chapter 8: The Normativity of Morality	197
Chapter 9: Objections and Implications	223
<i>Bibliography</i>	249

## DETAILED CONTENTS

---

<i>Abstract</i>	1
<i>Acknowledgements</i>	8
CHAPTER 1: MORAL COGNITIVISM	
1.0 The thesis	9
1.1 The open question argument	10
1.2 ‘The moral problem’	14
1.3 The practical reason approach	17
1.4 Moral disagreement	22
1.5 Other arguments for non-cognitivism	27
1.6 Conclusions	32
CHAPTER 2: MORAL JUDGEMENT INTERNALISM	
2.1. Introduction	33
2.1. The argument for internalism about moral judgements	33
2.2 Miller’s objection to Smith’s argument	36
2.3 Virtue and the motive of duty	38
2.4 Thoughts, desires, and dispositions	41
2.5 Moral rationalism	43
2.6 Internalism and the practical reason approach	47
2.7 Volitional impossibility	51
2.8 The motivational implications of normative considerations	52
2.9 Variable motivation	54
2.10 Amoralists and cynics	55
2.11 Internalism and Humeanism	57
2.12 Conclusions	60
CHAPTER 3: COMMONSENSE AND THE HUMEAN THEORY	
3.0 Introduction	61
3.1 The Humean theory: admittedly a dogma	61
3.2 Some Humean claims	63
3.3 Seven conceptions of desire	66
3.4 Commonsense and the Humean theory	68
3.5 The Humean theory: a ‘folk psychology’?	69
3.6 Commonsense explanations	72
3.7 Ordinary reasons explanations	75
3.8 Reasons explanations and cultural variation	77
3.9 Commonsense explanations in terms of reasons	79
3.10 Are belief-desire pairs sufficient to explain actions?	83
3.11 Norms and shared situations	87
3.12 Conclusions	90
CHAPTER 4: PLATTS’ DILEMMA	
4.0 Introduction	91
4.1 Hume and the Humean theory of motivation	91
4.2 Platts’ dilemma	97
4.3 Phenomenological and ordinary conceptions of desire	99
4.4 Pro attitudes and primary reasons	103
4.5 A constitutive explanation of action	109
4.6 Conclusions	113

## CHAPTER 5: SMITH’S ARGUMENT FOR THE HUMEAN THEORY

5.0 Introduction	114
5.1 Smith’s argument	114
5.2 The teleological conception of action	116
5.3 Beliefs about action	118
5.4 Belief and ‘direction of fit’	122
5.5 The dispositional conception of desire	124
5.6 Desire and ‘direction of fit’	129
5.7 Desires, dispositions and counterfactuals	131
5.8 Passive action	135
5.9 Conclusions	139

## CHAPTER 6: NORMATIVITY AND THE HUMEAN THEORY

6.0 Introduction	141
6.1 The normative reading of ‘direction of fit’	141
6.2 The assumption of means-end rationality	144
6.3 Motivating beliefs	147
6.4 Beliefs and desires, facts and values	152
6.5 Cognitivist internalism about reasons judgements	155
6.6 Nihilism and Humeanism	158
6.7 Three more Humean arguments	160
6.8 Conclusions	165

## CHAPTER 7: NORMATIVITY AND SUBJECTIVE MOTIVATIONS

7.0 Introduction	166
7.1 Why desired-based reasons are problematic for morality	166
7.2 Arguments for instrumentalism	168
7.3 Arguments against instrumentalism	171
7.4 Williams’ argument for internalism	174
7.5 The anticipated externalist objection	176
7.6 McDowell’s externalism	178
7.7 Smith’s objection to internalism	180
7.8 Internalism and vice	185
7.9 Internalism on epistemology and deliberation	186
7.10 The conditional fallacy	189
7.11 Internalism doesn’t satisfy the guidance condition	191
7.12 Conclusion	192

## CHAPTER 8: THE NORMATIVITY OF MORALITY

8.0 Introduction	194
8.1 The normative question	194
8.2 Realism about reasons	195
8.3 Normative regress	199
8.4 Universal reasons and the categorical imperative	201
8.5 Shared reasons	204
8.6 Maxims and the moral law	207
8.7 Second nature	211
8.8 Reasons and obligations on the practical reason approach	213
8.9 Internalism vindicated	216
8.10 Conclusions	219

## CHAPTER 9: OBJECTIONS AND IMPLICATIONS

9.0 Introduction	220
9.1 Williams’ objections	220
9.2 MacIntyre’s criticism of Kant	224

9.3 Cohen's objection to Korsgaard	228
9.4 Wallace's objection	229
9.5 Enoch's objection	232
9.6 The objection from particularism	236
9.7 Other objections	238
9.8 Cognitivism vindicated	241
9.9 Moral disagreement on the practical reason approach	243
9.10 Conclusions	244
<i>Bibliography</i>	245

### Statement of Copyright

*The copyright of this thesis rests with the author. No quotation from it should be published without the prior written consent and information derived from it should be acknowledged.*



## Acknowledgements

This thesis was written with the support of the ARHC and the Durham Doctoral Fellowship scheme. My supervisor, Geoffrey Scarre, has been a constant source of help and encouragement. I also received support from Simon James and Benedict Smith, who each acted as a second supervisor for part of the time I was writing the thesis. Benedict was very generous with his time, helping me to begin to formulate vague ideas into a coherent argument. Many other people at Durham University Philosophy Department have also helped me along the way, particularly Matthew Ratcliffe. It was Soren Reader's lectures that introduced me to ethics as a first year Biology student taking an elective in philosophy, and it was only the encouragement of my tutor, Bill Pollard, that led me to study philosophy beyond this introductory course.

Many of the philosophy postgraduates at Durham have helped me while I was writing this thesis. Comments and criticisms at the weekly *Eidos* seminars have helped me to develop some of my arguments and to avoid a number of errors that I might otherwise have made. Conversations with Emily Simon Thomas, Richard Stopford, Alex Carruth, Ian James Kidd, Liz McKinnell, Michael John Turp and Donnchadh O'Connaill helped me to clarify what I wanted to say. My sanity benefited greatly from my unspoken agreement with Heather Hei-Tai Yeung never to discuss academic work. Beth Hannon provided encouragement and a sympathetic ear; Arlette Frederik provided encouragement and tea. As well as talking amusing nonsense, Ian James Kidd helped by encouraging me to drink cheap wine in La Spaghetтата, by tolerating me writing in his living room in the evenings, and by providing tea and food.

The staff at the Durham branch of Caffè Nero tolerated my constant presence whilst the thesis was beginning to take shape. In the later stages, parts of chapters eight and nine were written in the Starbucks at the west end of Oxford Street and Costa Coffee at Mornington Crescent station. Sphongle, A. R. Rahman, M.I.A., Nitin Sawhney, Athlete, the Bluetones, Idlewild, Death Cab For Cutie, Elliot Smith, Ryan Adams, Chinty and Iraj provided the soundtrack to which the thesis was written. Michael Smith helped me to understand his account more clearly by asking me to explain where I thought he was wrong. Conversations with Ulrike Heuer and Benjamin Kiesewetter made me think more carefully about the claims I was making. Amanda Taylor provided useful comments on chapter three, and Donnchadh O'Connaill provided helpful feedback on chapters three and seven. Emily Simon Thomas, Ian James Kidd, Alex Carruth and Richard Stopford read other parts of the thesis and provided useful comments. The audiences at the 12<sup>th</sup> Durham-Bergen conference and the Edinburgh Epistemology workshop provided useful feedback on material that eventually found its way into chapters eight and nine. I am grateful to my parents, Michael and Carole Proctor, who encouraged and supported me throughout my studies.

## Chapter 1. Moral Cognitivism

### 1.0 The thesis

When we make moral judgements and act morally we recognise and respond to reasons that are there whether we recognise them or not. This is the claim defended in this thesis. It has two aspects. The first is that acts of moral judgement aspire, sometimes successfully, to moral knowledge. This is moral cognitivism. The second is that moral truths report reasons for action. In responding appropriately to these reasons we are motivated to action. Hence, it is claimed, there is a moral reality that we respond to in both cognition and action. In responding in these ways, we respond appropriately. The second aspect of this claim – that relating to motivation and action – might be referred to as moral internalism. I hesitate to use this term before discussing the matter in detail, since internalism can take many forms. Initially, then, I refer to the idea that morality gives us reasons for action, occasionally referring to this as ‘practical objectivity’ to contrast it with the ‘theoretical objectivity’ of moral cognitivism. However, I think of these as two aspects of the same central claim, linked by the idea of a practical reason. The moral truths that we can have knowledge of *are* the truths about the reasons for action that morality provides. Or so I shall be arguing in this thesis.

My central claim has the backing of our intuitions about morality; it is supposedly the commonsense view, reporting how morality ordinarily *seems*<sup>1</sup>. However, many metaethicists reject one or both of its two aspects. Those who reject cognitivism are either error theorists, who claim that moral judgements are universally false beliefs or non-cognitivists, who claim that moral judgements are not beliefs at all. In this chapter I focus on what I take to be the two strongest non-cognitivist arguments, the arguments from moral disagreement and practicality. I conclude that neither is conclusive but that both highlight challenges that cognitivists must meet. In particular, a tension between cognitivism and the practicality of morality arises in the light of the Humean theory of motivation when practical objectivity is formulated as ‘moral internalism’.

---

<sup>1</sup> See, for example, Blackburn 1998 p. 301; Crary 2002 pp. 373-4; Davidson 2004 p. 39; Lovibond 2002 pp. 3-4; Mackie 1977 pp. 30-5; McKeever and Ridge 2006 p. 113; Shafer-Landau 2006; Smith 1994 p. 1; Superson 2009 p. 3; Williams 1985 pp. 132-55

In chapter two I consider whether there is any good argument for moral internalism. I conclude that often arguments for internalism have failed and arguments for externalism have succeeded because internalism has been mischaracterised. I identify a form of internalism that corresponds to practical objectivity but which makes the task of arguing for internalism far harder than most internalists have imagined. Moreover, if the Humean theory is correct, this form of internalism is incompatible with cognitivism. However, I argue that we should reject the Humean theory. I argue that the Humean theory is not commonsense (chapter three), that it was not established by Hume (chapter four), that there is no good argument for the Humean theory, and that the Humean theory is false (chapters four to six). With the tension between cognitivism and practical objectivity removed, I argue for practical objectivity (chapters seven and eight). I then claim (chapter nine) that practical objectivity entails cognitivism (showing both error theory and non-cognitivism to be false), and that the form of cognitivism established is immune to the non-cognitivist argument from moral disagreement.

The case for cognitivism is compelling only if there are strong arguments for cognitivism *and* there are independent reasons to believe the arguments for non-cognitivism are unsuccessful. In this chapter I examine arguments for non-cognitivism. I identify the two arguments that were both the most influential in bringing about the non-cognitivist dominance of the 20<sup>th</sup> century and remain the strongest non-cognitivist arguments. The first is the open-question argument (§ 1.1) and modified forms thereof (§§ 1.2-1.3) and the second is the argument from moral disagreement (§ 1.5). I also briefly consider (§ 1.6) three less promising arguments for non-cognitivism sometimes offered alone or in combination with the open-question argument<sup>2</sup>. These are the arguments from naturalism, supervenience and the use of moral language to influence conduct.

## 1.1 The Open Question Argument

Moore's (1903) 'open-question argument' (OQA) was designed to support cognitivist non-naturalism. However, as we shall see, this argument later came to be wielded by

---

<sup>2</sup> I do not discuss Rorty's argument from the contingency of language. Rorty writes: "if the demands of morality are the demands of a language, and if languages are historical contingencies, rather than attempts to capture the true shape of the world or the self, then to 'stand unflinchingly for one's moral convictions' is a matter of identifying oneself with such a contingency" (1989 p. 60). The demands of morality are *not* the demands of a language. Moreover, one cannot show that something is contingent by showing that it requires a contingent language for its expression.

non-cognitivists against all forms of cognitivism, including non-naturalism. Here is the OQA<sup>3</sup>:

Moore's Premise: There is such a property as goodness

OQA Premise: Whatever natural property NP you fix on, it is an open question whether things that have NP also have goodness

OQA Conclusion: There is no natural property NP such that NP is identical with the property goodness

Moore's Conclusion: The property of goodness is a non-natural property.

If this argument is valid then either goodness is a non-natural property (Moore's Conclusion is true), or it is not a property at all (Moore's Premise is false). To many, Moore's non-naturalist account of goodness seemed metaphysically and epistemologically suspect, so they rejected Moore's Premise and embraced non-cognitivism. Let us look more closely at the OQA. Miller (2003 pp. 13-4) reconstructs the OQA roughly as follows:

- (1) If the predicate 'good' is synonymous with, or analytically equivalent to, the naturalistic predicate 'N' then it is part of the meaning of the claim that 'x is N' that 'x is good'.
- (2) If it is part of the meaning of the claim that 'x is N' that 'x is good' then someone who seriously asked 'Is an x which is N also good?' would betray some conceptual confusion.
- (3) Given any natural property N it is always an *open question* whether an x which is N is good. That is to say, it is always a *significant* question, of any x which is N, whether it is good: asking the question 'Is an x which is N also good?' betrays no conceptual confusion.

Therefore,

- (4) It cannot be the case that 'good' is synonymous with, or analytically equivalent to, 'N'.
- (5) The property of *being good* cannot as a matter of conceptual necessity be identical to the property of *being N*.

---

<sup>3</sup> As reconstructed by Thomson (2006 p. 233).

Miller objects that since we can distinguish between sense and reference (Frege 1892), it is possible that ‘good’ and predicate ‘N’ might have different senses yet refer to the same natural property, N. Property identities may be synthetic and *a posteriori* (Kripke 1980; Putnam 1975). This allows that the OQA works against analytic naturalism but not synthetic forms of naturalism such as Cornell Realism. Just as it was discovered that ‘water’ and ‘H<sub>2</sub>O’ refer to the same substance, we might discover that ‘good’ and (e.g.) ‘that which maximises happiness’ refer to the same property. Sayre-McCord (2008) writes:

Many argue that, while Moore’s argument shows that thinking something is good is different than thinking it is pleasant... it does not show that being good is different from being pleasant... After all, they point out, thinking that some liquid is H<sub>2</sub>O is different from thinking it is water and one might intelligibly wonder whether what one grants to be water really is H<sub>2</sub>O. Yet that does not show that being water is anything other than being H<sub>2</sub>O.

Elstein (2006) follows Horgan and Timmons (1992) in claiming that this objection fails to damage the OQA. We establish that water is H<sub>2</sub>O by investigating the structure of the watery stuff around here, where by ‘watery’ we refer to the natural features of water that facilitate our everyday identifications of water. It is analytic that water is watery.

By analogy, showing that good displays synthetic *a posteriori* identity to some natural property requires picking out those natural features of good things that facilitate our recognition of them as good. Call these good making features ‘G-features’. In the case of water, once we picked out the watery things (i.e. samples of water), we looked at their structure (and discovered it was H<sub>2</sub>O), so by analogy, once we have picked out the things with G-features we need to look at their structure. Then, according to the naturalist, we will discover some natural property, N, that all things with G-features share, and this property N is identical with goodness. However, for this account to work, and for the analogy with water to hold, it must be analytic that good things have G-features. Since G-features are natural properties, it is an open question whether something that has G-features is good. Therefore, ‘good’ cannot mean or be analytically equivalent to ‘having G-features’, and the identity between G-features and natural property N cannot therefore show that goodness is identical with N. Hence, if the original OQA works, it works against synthetic *a posteriori* naturalist claims as well as analytic naturalism.

This response apparently succeeds in defending the OQA<sup>4</sup>. However, there is another problem for the OQA. It seems to rely on the assumption that there are no interesting philosophical analyses of concepts and that any correct analysis of a concept must be one that would be immediately recognised as correct by anyone conceptually competent. Competence with a concept consists of knowing how to employ the concept, and this does not presuppose the propositional knowledge describing that ability. Therefore, one might be competent with the concept ‘good’, yet when offered a correct analysis of ‘good’ one may fail to recognise it as such (Miller 2003 pp. 16-7).

There is a further reason for thinking that we should reject Moore’s Premise independently of whether we think Moore’s Conclusion is unsatisfactory. The predicate ‘is good’ is semantically incomplete (Geach 1956; Ziff 1960; von Wright 1963; Foot 1985; Thomson 1997, 2006). Thomson writes:

The sentence ‘A is a red apple’ entails ‘A is red and A is an apple,’ for ‘A is red’ (like ‘A is an apple’) is semantically complete. By contrast, ‘A is a good dancer’ does not entail ‘A is good and A is a dancer’ – else ‘A is a good dancer and a bad tennis player’ would entail that ‘A is good and A is bad and A is a dancer and A is a tennis player’ (2006 p. 246).

Thomson claims that therefore when we ask ‘which things possess the property ‘goodness’?’, “taken literally, the only possible answer is ‘None’” (2006 p. 246). It is unclear whether Thomson’s observations apply across all uses of good – whether they apply to intrinsic goodness and moral goodness (Zimmerman 2001 p. 24). Moreover, it is not clear what the significance of Thomson’s observations would be if we pass the buck in the opposite direction to Moore and decide to understand statements about goodness as reporting facts about what we have most reason to protect or bring about (Scanlon 1998). Importantly, however, Thomson’s insights do not support non-cognitivism. That ‘A is good’ is semantically incomplete and strictly speaking cannot be true or false does not show that ‘A is a good pen’, for example, cannot be true or false.

These considerations can be applied to ‘rightness’ in a similar way: ‘Φ-ing is the right way to draw a boat’ does not mean that ‘Φ-ing is right and Φ-ing is the way to draw a boat’. If it did then ‘Φ-ing is the right way to draw a boat but the wrong way to

---

<sup>4</sup> However, it is unclear whether it is a decisive argument against Cornell Realism. For one thing, Cornell Realism is supposedly a non-reductive theory, denying that there analytic relationships between moral and non-moral terms and denying “that moral properties are identical to non-moral properties as a matter of synthetic fact” (Miller 2003 p. 139).

draw a cat' would mean 'Φ-ing is right and Φ-ing is wrong and Φ-ing is a way to draw a boat and Φ-ing is a way to draw a cat'. However, it does not follow that claims of the form 'Φ-ing is morally right' lack truth-values. We can note that 'Φ-ing is the right way to be moral but the wrong way to increase one's pension contributions' does not mean 'Φ-ing is right and Φ-ing is wrong and Φ-ing is a way to be moral and Φ-ing is a way to increase one's pension contributions.' Yet just as there are truths about how to increase one's pension contributions there might be truths about how to be moral.

## 1.2 'The moral problem'

The connection between the OQA and non-cognitivism is found in that which is easiest for non-cognitivism to account for and most problematic for Moore: the idea that morality is practical. As MacIntyre says, "Moore's account leaves it entirely unexplained and inexplicable why something's being good should ever furnish us with a reason for action" (1967 p. 244). Hare puts the point like this:

If we were to ask of a person 'What are his moral principles?' the way in which we could be most sure of a true answer would be by studying what he *did*. (1952 p. 1)

How does this relate to the OQA? Stratton-Lake and Hooker explain:

Non-cognitivists hold that the openness of the open question is explained by the possibility of motivational indifference to anything characterized in purely naturalistic terms... judging that something is good is *per se* to be motivated to act in certain ways towards that thing. But judging that something has some naturalistic property does not necessarily imply the presence of such a motivational attitude towards that thing (2006 p. 151)

Elstein (2006), a contemporary defender of Hare, expresses his understanding of the OQA by saying that moral questions are *substantive* and *practical*. Doubting whether producing happiness is right is an intelligible doubt that is not addressed by accusing the doubter of ignorance about the meaning of 'right'. Such doubts should be characterised as doubts about what one *should do* or, more directly, about *what to do*. Our linguistic practice with respect to 'good' is essentially linked to choice; "we should not speak of good billiard-cues, unless sometimes we had to choose one billiard-cue in preference to another" (Hare 1952 p. 128). Hence, we can see why Darwall, Gibbard and Railton say

that the OQA “came to bite the hand that first fed it, and, eventually, to count Intuitionism among its victims” (1992 p. 118). If moral judgements are to connect with action they must be able to *motivate* agents. However, “it appears no easier to see how an appropriate link to motivation or action could be logically secured if we were to substitute ‘*sui generis*, simple, nonnatural property Q’ for ‘naturalistic property R’” (*ibid.*). A modified OQA can therefore threaten any form of cognitivism:

1. Moral cognitivism: moral knowledge is possible. Therefore, moral judgements can be true. Moral judgements are judgements about moral properties.
2. Moral internalism: There is a conceptual or internal link between making a moral judgement and being motivated, *ceteris paribus*, to act as that judgement prescribes
3. Modified OQA: For any property, natural or non-natural, Q, it is possible that an agent judge that Q obtains but who fail to see any reason or have any motivation to act in accordance with that judgement<sup>5</sup>

Whilst rather different to the original OQA, this captures a thought implicit in the original argument that is drawn out by non-cognitivist uses of it. The thought is that the reason why calling something ‘good’ is quite different to predicating any natural (or ‘metaphysical’) property of it is that the former is necessarily linked to action in a way that the latter is not. Thus, a modified version of the OQA shows us why the OQA might seem to favour non-cognitivism: “the acceptance of internalism seems to be one of the major forces driving us towards non-cognitivism” (Mason 2008 p. 135). However, what begins to emerge is not so much an argument against cognitivism, but a metaethical puzzle that could be resolved in a number of ways: this heavily modified version of the OQA resembles what Smith has called ‘The Moral Problem’.

‘The Moral Problem’ is an apparent inconsistency between the objectivity of morality, the practicality of morality and the nature of agency. It can be stated in terms of “three apparently inconsistent propositions” (Smith 1994 p. 12):

---

<sup>5</sup> This is a modified version of a revised OQA stated by Miller (2003 pp. 34-5).



1. Moral cognitivism: Moral judgements of the form ‘It is right that I  $\Phi$ ’ express a subject’s beliefs about an objective matter of fact, a fact about what it is right for her to do.
2. Moral internalism: If someone judges that it is right that she  $\Phi$ s then, *ceteris paribus*, she is motivated to  $\Phi$ .
3. The Humean theory of motivation: An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume’s terms, distinct existences (*ibid.*).

There are two interesting things to note. Firstly, internalism appears in this inference and the modified OQA above. Whilst we undeniably have strong intuitions about morality being *practical* it is not immediately clear whether these are captured by internalism. It is not clear whether internalism is the same as what I have been calling ‘practical objectivity’, the claim the morality gives us reasons for action. Until more is said about these matters we cannot be sure whether either of these inferences undermines cognitivism.

Secondly, ‘The Moral Problem’ differs significantly from the modified OQA inference since the Humean theory of motivation appears in place of the modified OQA. Whilst the modified OQA seems highly plausible, we have not yet said anything about the Humean theory. The Humean theory is typically expressed in one of two ways, sometimes taken to be equivalent. The first is that reason alone cannot motivate. The second is that any action (or any action done for a reason) is to be explained by a desire of the agent and a belief of the agent about how to satisfy that desire. This belief-desire pair constitutes the motivating reason for the action.

Cognitivism entails that moral judgements are beliefs and internalism entails that moral judgements are intrinsically motivating; together they entail that some beliefs (moral beliefs) are intrinsically motivating. However, according to the Humean theory, no belief is intrinsically motivating; a belief can only motivate in combination with a suitably related desire. Hence, cognitivism, internalism and the Humean theory appear to form an inconsistent triad. Smith suggests that this inconsistency is merely apparent and that all three propositions can and should be held simultaneously. I disagree. I will argue that we should accept cognitivism, accept internalism only in a significantly modified form and reject the Humean theory altogether. However, I propose to defend cognitivism *via* defending a modified form of internalism, namely the claim that

morality always provides us with practical reasons. I adopt a practical reason approach according to which moral knowledge is made possible by the fact that morality gives us reasons for action. I now give a preliminary outline of this approach.

### 1.3 The practical reason approach

The claim that morality always provides us with practical reasons is sometimes expressed as the idea that moral requirements are categorical imperatives<sup>6</sup>. Speaking of categorical imperatives, Mackie says:

So far as ethics is concerned, my thesis that there are no objective values is specifically the denial that any such categorical imperative element is objectively valid. The objective values which I am denying would be action-guiding absolutely, not contingently ... upon the agent's desires and inclinations (1977 p.29)

However, there are two senses of 'categorical imperative' that Mackie does not quite keep apart. Foot explains the first:

The club secretary who has told a member that he should not bring ladies into the smoking room does not say, 'Sorry, I was mistaken' when informed that this

---

<sup>6</sup> It is also sometimes expressed by saying moral requirements are rational requirements (e.g. Smith 1994 pp. 84-91). This is confusing because on some understandings of rationality, it might be untrue that one ought always to do what is most rational. For example, being 'rational' is sometimes equated with maximising expected utility or fulfilling one's desires. In this sense, it is an open question as to whether one should do what is rational. Whilst rationality is taken by some (e.g. Parfit 1997 p. 99) to consist in responsiveness to reasons, others claim that rationality is an independent source of requirements (Broome 2000, 2004, 2007a p. 374;). Broome claims, "rationality may bring you to do things that you have no reason to do" (2000 p. 99). Broome thinks of rationality as a set of requirements applied to relations between states such as beliefs and intentions (2007b p. 161) such as the requirement to conform one's beliefs to *modus ponens*. Broome is agnostic on the question of whether rational requirements are normative. Should we share Broome's agnosticism? Consider *modus ponens*. As Broome says, "rationality requires of you that, if you believe it is raining and you believe that if it is raining the snow will melt, you believe the snow will melt" (2007b p. 168). To believe *not q* whilst believing *p* and *if p then q* is to believe the world is a way that it cannot possibly be. However, when one believes *p* and *if p then q*, *modus ponens* does not straightforwardly ground a requirement to believe *q*. It grounds a requirement to either believe *q* or to cease believing at least one of *p* and *if p then q*. We can understand conforming to *modus ponens* in terms of responding to reasons. There is a reason not to believe *not q* whilst believing *p* and *if p then q*: namely that these three propositions cannot be true together. This is only a *pro tanto* reason, and might be outweighed by other reasons. However, even if having a pair of contrary beliefs would prevent a war, "there would still be a good reason for me not to believe both of them, which is that no pair of contraries can be true" (Dancy 2004 p. 37). Hence, Broome's worries are discharged when we bear in mind that the reasons that rational requirements report are contributory reasons, not overall verdicts on what we should do. We can hold, as Raz does, that "[t]he normativity of all that is normative consists in the way it is, or provides, or is otherwise related to reasons" (2000 p. 34). However, I speak of moral requirements as giving us reasons both to avoid confusion and because in fact I do not think that the normativity of moral requirements is generally provided by their being (or being relating to) rational requirements.

member is resigning tomorrow and cares nothing about his reputation in the club. Lacking a connection with the agent's desires or interests, this 'should' does not stand 'unsupported and in need of support'; it requires only the backing of the rule. The use of 'should' is therefore 'non-hypothetical' in the sense defined. (1972 pp. 308-9)

Although etiquette or the rules of a club apply irrespective of an agent's desires, this makes them categorically binding only in a restricted sense:

[O]ne may reasonably ask why anyone should bother about what should ... from the point of view of etiquette ... be done... Considerations of etiquette do not have any automatic reason giving force, and a man might be right if he denied that he had reason to do 'what's done.' (Foot 1972 p. 309)

Likewise, the rules of a club are not necessarily reason-giving. This is not only because the rules of the club can fail to apply to you (e.g. because you are not a member) but also because they can apply without having any reason-giving force. When moral requirements are said to be categorical, they are contrasted with club rules and etiquette in this sense, since "it is supposed that moral considerations necessarily give reasons for acting to any man" (*ibid.*). The claim that moral requirements are categorical is therefore the claim that they always provide us with reasons for action. We might refer to this claim as 'practical objectivity' and to moral cognitivism as 'theoretical objectivity' to indicate that in bringing them together we are effecting "the reconciliation of morality's intellectual and practical dimensions" (Brink 1997 p. 4). Importantly, 'practical objectivity' is not the same as 'moral internalism'. If moral requirements are practically objective then they give us *normative* reasons for action. However, moral internalism, as stated above, is the claim that *making* or *holding* a moral judgement gives us a *motivation* or *motivating* reason for action<sup>7</sup>.

We are now close to seeing why the claim that moral requirements are categorical imperatives is closely linked to cognitivism. Dancy offers us the following piece of practical reasoning (2004 p. 38):

---

<sup>7</sup> We must distinguish between motivating and normative reasons. Dancy writes: "When someone does something, there will (normally) be some considerations in the light of which he acted... This issue about his reasons for doing it is a matter of motivation. There is also the question of whether there was a good reason to act in that way, as we say, any reason for doing it at all, one perhaps that made it sensible in the circumstances, morally required, or in some other way to be recommended... This second question raises a normative issue." (2000a pp. 1-2) Dancy says that there are not two *kinds* of reason, but "two contexts

1. I promised to do it.
2. My promise was not given under duress.
3. I am able to do it.
4. There is no greater reason not to do it.
5. So: I do it

He immediately notes that “there is a similar train of thought that ends not with (5) as presented, but with: 5\*. So I ought to do it” (*ibid.*). According to Dancy, the original reasoning concerns the favouring relation, whilst the modified reasoning concerns the right-making relation. The original train of thought moves from the recognition of a reason to perform an action,  $\Phi$ , to actually  $\Phi$ -ing because one sees that overall there is most reason to  $\Phi$ . Supposing such deliberation is rational and that (1) is a genuine reason for action, action on this basis is an instance of the ‘practical objectivity’ of morality. An agent who acts on this basis responds to a reason-giving moral consideration and hence responds to the categorical (i.e. reason-giving, normative) nature of morality.

Following the modified train of thought (to 5\*), we move from thinking that there is a reason to  $\Phi$  to judging that the situation requires one to  $\Phi$ . Since this is a judgement about what the normative reasons require when taken together, it is truth apt if two conditions are met. Firstly, it must be possible that normative reasons bear upon this case<sup>8</sup>. We assumed this in the original progression. Secondly, it must be possible to move from judgements about individual normative reasons to judgements about overall requirements in a justified manner. This second point seems to be vindicated immediately just by looking at the piece of practical reasoning in question. If (1) is a normative reason then (2) and (3) are enablers for that reason and (4) enables a justified move to (5) (cf. Dancy 2004 pp. 40-1).

If there are genuine practical reasons and valid ways of judging what they require when taken together, we end up with truths about what one ought to do. Supposing that some normative reasons are moral considerations (we also assumed this in the original progression), and that one can make valid judgements about what the

---

in which we use the notion of a reason” (2000a p. 20). Hence, “one and the same reason can be both motivating and normative” (2000a p. 6).

<sup>8</sup> That is to say, deciding to act in a particular way must be the *kind of thing* that there can possibly be reasons for and against. This seems indisputable, because it seems that anything that is under the control

moral reasons require when taken together, we end up with truths about what one ought to do, morally speaking. This is the ‘theoretical objectivity’ of morality, or moral cognitivism: there are truths about what morality requires and acts of judgement through which we can access those truths. Since all of the assumptions required to yield theoretical objectivity were also required to yield practical objectivity, in vindicating practical objectivity we vindicate theoretical objectivity<sup>9</sup>. To think in this way is to think of philosophical questions about morality as being *addressed to us* rather than being *about us*. Then, when we reflect on moral principles, we will see that “the philosophical question about them is not so much how we know them as why we have to conform to them. Or rather, those two questions become one. In answering the second we will already have answered the first” (Korsgaard 1996a xii).

This practical reason account also allows us to provide a cognitivist account of the meaning of ‘good’ that is not Moorean and yet does not commit us to holding that goodness is identical with any natural property. We can adopt a modified version of the ‘buck-passing’ account of goodness developed by Stratton-Lake and Hooker (2006) and Scanlon (1998) as an alternative to the Moorean view. According to the buck-passing view, “to be good is to have properties that provide reasons to care” (Stratton-Lake and Hooker 2006 p. 165) and “the fact that A has such-and-such natural properties *provides* the reasons we have” (p. 154). Goodness supervenes on these natural properties. However:

[T]hinking of the goodness of A as adding to the reasons provided by the base properties of A commits a kind of category mistake. If the reason why the holiday resort is good is because it is pleasant, then the reason to choose the holiday resort is just that it is pleasant, not because it is pleasant and because it is *good* (*ibid.*).

According to the Moorean view, “[t]he base properties of A make A good, and the goodness of A gives us a reason to care about A. Reasons supervene on goodness, and goodness supervenes on the base properties” (Stratton-Lake and Hooker 2006 p. 155). The Moorean does not claim that goodness gives us an *additional* reason to care about

---

of beings who are capable of reflecting on their activity and thinking in terms of reasons is something that it is possible for there to be reasons for and against.

<sup>9</sup> Smith accepts a similar view. He writes: “when we do not know whether such [i.e. moral] normative reasons exist or not, we have no choice but to suppose that they are the sorts of things that we are looking for in looking for moral facts.” (1997 p. 117)

A; they claim that the only reasons we have to care about A are based on the goodness of A. Nonetheless, Scanlon's view is preferable to the Moorean account:

The strong intuition Scanlon points to is that we seem to have exhausted the list of reasons to care about something once we have listed the features that make it good. We do not need to add the fact that it is good to complete the list. This intuition clearly assumes that the good-making features are reason-giving, which the Moorean view simply denies (*ibid.*).

This account seems plausible subject to a couple of modifications. Firstly, whilst it seems right to understand 'good' as a property supervenient on reason-giving features, these features are not always reasons to *care*. Hare's analysis brings out this fact: a good motorcar is one with properties that give us reason to *choose* it over other motorcars, not one with properties that give us reasons to care. There are a variety of responses that good-making features can make appropriate, and hence what these features give us reason to do is diverse – it is not limited to caring.

Secondly, Stratton-Lake and Hooker hold that "reasons and goodness always share the same ground" (2006 p. 159). This seems to obliterate Ross' (1930) distinction between the good and the right. We can account for goodness in terms of reasons without having to hold that reasons *always* ground goodness. Sometimes reasons for action contribute towards making an action *right* or *appropriate* but not towards making it good. Subject to these modifications, this account allows us to hold that thin evaluative concepts such as 'good' and 'right' are primarily evaluative and normative whilst pointing us towards a way of maintaining that judgements of rightness and goodness can report or constitute moral knowledge.

A point of clarification is required. I will argue that moral requirements always provide reasons for action. However, it is sometimes claimed that moral requirements are overriding in the sense that, when they are in play, they always outweigh all other considerations. In Superson's view, defending this claim is necessary to overcome moral scepticism. She writes:

[W]e do worry that skeptics about acting morally will treat others badly, for if moral reasons do not override other reasons for action, then rational persons will do well not to act on them. (2009 p. 4)

Superson's claim does not seem quite right<sup>10</sup>. Whether people who respond appropriately to reasons act morally is not crucially dependent on whether or not moral reasons are, or are taken to be, overriding. Non-overriding reasons must be taken into account in deliberation. Perhaps Superson is worried that unless moral reasons are overriding, non-moral reasons will tend to outweigh them in the deliberations of rational, reason-responsive agents, perhaps because selfish concerns are stronger and more abundant than altruistic ones. There are no grounds for this worry. Moral reasons might be *weighty* without being overriding. Ordinarily we do generally consider moral considerations as being among the most important kinds of demand on our conduct without considering them overriding. Importance is distinct from overridingness. If moral demands were overriding they would act like trumps in a game of cards – the least moral demand would outweigh even the greatest conjunction of non-moral demands. This would make moral reasons intrinsically decisive rather than contributory (Dancy 2004 p. 43). One would be obliged to sacrifice one's most important projects in order to keep even the most trivial promise. This claim is implausible; failing to defend it involves no form of moral scepticism. It is enough to show that moral requirements are important and reason-giving.

## 1.4 Moral disagreement

Moore's non-naturalism, in relying on a simple intuitionist epistemology, lacked a plausible account of moral disagreement. This explains why early non-cognitivists made moral disagreement a central feature of their theories or even prerequisite for any successful theory (see Stevenson 1937 pp.16-8, 1944 ch.1; MacIntyre 1967 p. 250). Moral disagreement is often seen as threatening not only intuitionism but all forms of cognitivism. There are a number of potential arguments, some targeting the truth-aptness of moral judgements and others targeting the possibility of providing sufficient epistemic justification to allow moral knowledge. I discuss (1) the argument from shared concepts, (2) the argument from no method and no progress, (3) the existence of

---

<sup>10</sup> If Superson replaced 'will' with 'may' and added 'on some occasions', her claim would be more plausible. This points to a legitimate worry that (e.g.) selfish reasons might often end up outweighing moral reasons, leading rational people to act rather selfishly on the whole. However, I think this worry will be dispelled if my arguments for the importance of moral reasons are successful (chapter eight) and if we resist the temptation to think that an agent's own desires are the main source of their reasons (see chapter seven).

rationally irresolvable disagreement, (4) the epistemological challenge and (5) inference to the best explanation.

Firstly, then, the argument from shared concepts. Hare (1952 pp. 148-9) asks us to imagine a Christian missionary visiting a cannibal island. By coincidence the cannibal word for 'good' is also 'good'. If the Christian and the cannibals both use the word 'good' evaluatively, they can communicate about morals. When the missionary calls meek and gentle people 'good' and the cannibals call those who collect a lot of scalps 'good' they will be genuinely disagreeing. However, if 'good' had a descriptive meaning it would have different descriptive meanings in English and in the cannibal language. To change the example, imagine the cannibal word for good was 'cood'. If 'good' and 'cood' are used evaluatively, 'x is good but not cood' is a contradiction. Hence, a disagreement arises when one says 'x is good' and another says 'x is not cood'. However, if 'good' and 'cood' are used descriptively then 'good' would mean "among other things 'doing no murder'" (Hare 1952 p. 149) and 'cood' would mean "among other things 'productive of maximum scalps'" (*ibid.*). The upshot of this would be that no claim about 'goodness' could contradict a claim about 'coodness', and hence the words 'good' and 'cood' could not be used to disagree about morals.

The point is that when English speakers disagree about whether something is good they are not disagreeing about the meaning of the English word 'good'. They are engaged in moral disagreement, not linguistic disagreement. Hare's cannibals and missionaries argument might appear to be a version of the OQA, since it seems closely related to the idea that it is an open question whether 'production of maximum scalps' is good. However, the argument is sufficiently distinct to constitute an independent challenge to cognitivism. Its distinctness lies in the appeal to intuitions about moral disagreements, to the idea that intuitively, moral disagreements are typically substantive, not conceptual. However, we can agree with non-cognitivists in accepting this claim whilst retaining a cognitivism based on the practical reason approach. According to this approach, moral disagreements are substantive since participants who resolve a moral disagreement necessarily change some of their views about what they have reason to do. If judgements about the goodness of something, X, are justified by reasons that we have to (e.g.) choose X then they are substantive disagreements, since they are not only disagreements about the validity of the goodness judgement but also



disagreements about what there is reason to choose. The argument from shared concepts is no threat to the practical reason approach.

Secondly, there is the argument from no method and no progress. Whilst disagreement in other domains (e.g. science) is not typically taken to undermine cognitivism, moral cognitivism might be undermined by the *lack of a method* for resolving moral disagreement. Moreover, our views on morality seem not to be progressing towards agreement (Enoch 2009 p. 29; cf. Williams 1985 ch. 8). Enoch writes:

Moral disagreement runs much deeper than disagreement in the sciences, because typically, or at least often, in cases of moral disagreement there is no deeper agreement underlying it, no agreement about how to settle the more superficial disagreement. (2009 p. 34)

The claim that there is no method for deciding cases of moral disagreement is open to a number of interpretations. Enoch suggests the following possibilities:

1. We literally have no way of proceeding when faced with moral disagreement (2009 p. 35)
2. “[T]hough we do proceed in any number of ways, none of them is justified” (*ibid.*)
3. “[T]here is no method that is guaranteed to lead to agreement” (*ibid.*)
4. “[T]here is no method of settling the disagreement that is itself accepted as the proper method by all parties to the original disagreement” (*ibid.*)

The first of these is false:

We do proceed in any number of ways both in conversation and deliberation, and in action... We try to reason, to convince, to draw analogies and make comparisons, to reduce ad absurdum, to draw conceptual distinctions, to imagine what it would be like to be on the other side, to engage each others’ emotions and desires, to rely on authority, and so on. (Enoch 2009 p. 36)

The second claim is true only if all the methods we use for resolving moral disagreement, including those just mentioned, are unjustified. This seems unlikely. Intuitively, *all* the methods just mentioned would seem to be justified at least in some cases. Even relying on authority might be justifiable if the authority is a legitimate one.

The onus is therefore on the non-cognitivist to convince us otherwise. The third possibility (that there is no method *guaranteed* to lead to moral agreement) seems unlikely to support non-cognitivism. There are many reasons why we might not be able to convince someone of something that is true: stubbornness, irrationality, errors in background beliefs, and limitations in time, for example. The fourth and last possibility is that since there is disagreement about the method for resolving moral disagreements, there are no moral truths. We have no reason to accept this inference. The fact that people cannot agree on a method for resolving moral disagreements does not show that there is no such method – there are such methods, as we have already said, and some of them might well be justified. Therefore, the argument from no method and no progress does not threaten cognitivism.

Thirdly, if there are actual moral disagreements that are not rationally resolvable, this might undermine cognitivism. This is closely related to the idea that none of the methods we use to resolve disagreement are justified, but it extends further to take in any possible rationally justified method, not just those that we actually ordinarily employ. Gowans says that it appears, “that there are widespread and deep moral disagreements that appear persistently resistant to rational resolution” (2000 p. 2). Gowans thinks that this ‘Disagreement Thesis’ renders moral cognitivism problematic (2000 p. 3). If moral disagreements resist rational resolution, either moral judgements are not beliefs or moral judgements are beliefs that cannot be adequately justified. If moral judgements were beliefs that could be adequately justified, presenting those justifications would convince all rational people, and moral disagreements would be rationally resolved.

However, this does not undermine moral cognitivism on the practical reason approach. To see this we only need to ask whether recognition of reasons is guaranteed by rationality. Usually being rational is characterised fairly minimally to involve things such as consistency between mental states, giving proper weight to evidence and acting on intentions and recognised practical reasons. Since recognition of reasons requires intellectual and practical virtues<sup>11</sup>, it is not guaranteed by rationality. However, if we expand our notion of rationality to include everything that could guarantee recognition

---

<sup>11</sup> E.g. a good detective (a) notices features that constitute reasons for belief and (b) recognises their significance. The features of her that allow her to display this excellence that other people lack are intellectual virtues. Likewise, a caring person notices the needs of others because they notice features that others ignore and because they recognise their significance. These interdependent factors partially

of reasons, the Disagreement Thesis becomes incredibly implausible. Even if there are any perfectly virtuous agents, we have yet to see any evidence that those involved in interminable moral disagreements with each other are to be counted among their number.

A fourth argument against cognitivism from moral disagreement is based on the fact that cognitivists need to give a moral epistemology, whereas non-cognitivists do not. Moral disagreement provides a direct argument against any epistemology that relies on self-evidence. However, there are plenty of other options for cognitivists: a practical reason approach is one. Even so, moral disagreement might apply more pressure on the cognitivist epistemology than this:

Perhaps the problem moral disagreement points to is that in cases of moral disagreement... we have no way of knowing, of finding out who is right and who wrong. (Enoch 2009 p. 45)

However, this does not threaten the claim that moral knowledge is possible. That moral knowledge is possible does not entail that it must be possible to be able to provide universally accepted accounts that show which moral claims are true and false and which cases are instances of moral knowledge. An argument for cognitivism *might* show this, but it does not have to. It need only show that e.g. we can *know* when something is morally required of us, not that we can *know that we know* when something is morally required. Therefore moral disagreement once more provides a challenge for cognitivism, but not an argument against it.

A fifth argument against cognitivism suggests that moral disagreement is better explained if we reject objectivity than if we accept it. Mackie writes:

[T]he actual variations in the moral codes are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values. (1977 p. 37)

Subjectivism, relativism and non-cognitivism predict and explain wide-ranging disagreement and demand to know, “if there is an objective, universal, moral truth, why is it hidden from so many people in so many matters?” (Enoch 2009 p. 21) A familiar

---

constitute this practical virtue. These examples illustrate how recognition of reasons for action is often dependent on both intellectual and practical virtues.

cognitivist response is to suggest that much apparent moral disagreement is not essentially moral, but derives from other forms of disagreement, such as metaphysical views about life after death. Whilst this may describe many cases of moral disagreement, it is implausible that all moral disagreement can be explained in this way. Therefore cognitivists need to add something else to their account.

In responding to these worries we do well to point out that “[i]t is not that disagreement needs explanation and agreement does not, but that in different contexts disagreement requires different sorts of explanation, and so does agreement” (Williams 1985 p. 133). Perhaps then, when it comes to morality, “given our cognitive shortcomings agreement rather than disagreement is what calls for explanation” (Enoch 2009 p. 23). Non-cognitivists will be quick to point out that our cognitive shortcomings do not preclude agreement in other areas. Nonetheless, in the case of moral disagreement there are a number of additional factors that help to explain why disagreement is so persistent. Moral complexity, limits in empathy and emotional sensitivity, the influence of prejudice, self-interest and other commitments, manipulation by others and the existence of indeterminate moral issues are some of the most prominent (Enoch 2009 p. 25). Enoch highlights self interest as a particularly powerful distorting factor on moral belief: since often so much is at stake, we may refuse to acknowledge moral truths that we dislike. Enoch says: “self interest can thus serve to explain not just the scope of moral disagreement, but also the difference between the scope of moral and other disagreement” (2009 p. 26; cf. Nagel 1986 p. 148; Shafer-Landau 2006 p. 219). Since typically little is at stake personally when it comes to difficult questions in mathematics, for example, self interest has little distorting effect in that domain. Hence, the cognitivist has plenty of explanatory resources with which to fend off the non-cognitivist argument to the best explanation.

## 1.5 Other arguments for non-cognitivism

Naturalism is thought to constitute an advantage for non-cognitivists. Suppose for the sake of argument that naturalism is true, and that naturalism is the claim that the only properties and entities that exist are natural properties and entities. Suppose that natural properties and entities are those that are studied by the natural sciences and that enter into causal relations or can be detected by the senses (Miller 2003 p. 11). Non-cognitivists demand to know what cognitivists say about moral properties. To say that

moral properties are non-natural properties entails that naturalism is false. However, we are supposing that naturalism is true. The other option for the cognitivist is to say that moral properties are natural properties or are reducible to natural properties. However, the modified OQA shows us that neither of these forms of naturalist cognitivism is viable. Natural properties are not motivating or reason-giving; they do not direct us to action. Moral properties are and do. The two cannot be identical.

Naturalism seems problematic for cognitivism but it has no problematic consequences for non-cognitivism since non-cognitivists do not posit any moral properties. One consequence of this is that non-cognitivism has the advantage of “metaphysical and epistemological economy” (Miller 2003 p. 53). However, I do not think that this establishes an advantage for non-cognitivism. The first reason is that metaphysical economy should not be sought at the expense of failing to account adequately for the phenomenology of the moral life. Blackburn, a proponent of non-cognitivism, says that his non-cognitivist theory “intends to ask no more from the world than what we know is there” (1984 p. 182). However, he also writes “our first-order ethical practice is based on the presupposition that there are objective, independent, binding ethical facts, facts that exert pressure on all rational beings, whereas in truth there are no such facts” (1998 p. 301). The disadvantage of denying an assumption on which all our moral practice is based seems to far outweigh any advantage gained in metaphysical economy. Moreover, asking ‘no more from the world than what we know is there’ is a virtue of non-cognitivism but not of cognitivism only on the question-begging assumption that we lack moral knowledge. If we have moral knowledge then we *do* know that moral properties are ‘there’.

Another weakness of the argument from naturalism is that it is not clear whether naturalism is true or even exactly what the naturalist claim consists of. The question of what makes a property natural is only solved by reference to the subject matter of the natural sciences if we know what makes a science ‘natural’. The appeal to causality and the senses is not much better. For one thing, causality is notoriously difficult to account for philosophically. For another, the predicate ‘natural’ is not supposed to apply to supernatural beings such as God, but God is not usually thought of as being causally inert. Furthermore, if we take this naturalist project seriously, we would have to doubt the existence of properties such as ‘validity’ and ‘truth’ since neither the validity of an argument nor the truth of a proposition is something perceived by the senses. There are two conclusions that I think we should draw. Firstly, the natural/non-natural distinction

is not particularly useful in this context. Secondly, if naturalists do insist upon drawing this distinction it should not worry us. Until they come up with a clear basis for the natural/non-natural distinction and compelling reasons for rejecting the existence of anything non-natural there is little reason to worry if moral properties fall on the wrong (i.e. non-natural) side of it. Therefore, as things stand, naturalism does nothing to support non-cognitivism.

Supervenience is sometimes claimed to support non-cognitivism. Supervenience can be considered to be an insight about moral properties or moral assertions. Cognitivists might think of moral supervenience in terms of moral properties supervening on natural properties, and this just means that whenever a moral property is present it is present *because* of the natural properties that are present (Zangwill 2008c). As Miller puts it: “if two things have exactly the same natural properties, then they also have the same moral properties. If you find that two things have different moral properties, you must also find that they differ in some way in respect of their natural properties” (2003 p. 31). Importantly, to claim that A-features supervene on B-features, or A-talk supervenes on B-talk, is to make the modal claim that there *cannot* be a difference in A without a difference in B (McLaughlin and Bennett 2010). Hare, who first used the term ‘supervenience’ in this context, explains it in a way that is neutral between non-cognitivist and cognitivist interpretations:

[L]et us take that characteristic of ‘good’ which has been called its supervenience. Suppose that we say ‘St. Francis was a good man’. It is logically impossible to say this and to maintain at the same time that there might have been another man placed in precisely the same circumstances as St. Francis, and who behaved in them in exactly the same way, but who differed from St. Francis in this respect only, that he was not a good man. I am assuming, of course, that the judgement is made in both cases upon the whole life of the subject, ‘inner’ and overt. (Hare 1952 p. 145 cf. pp. 80-1)

Hare thinks that this feature of supervenience is to be explained by the fact that the characteristics of the man form the minor premise of an inference in which a standard for judging men morally is the major premise and the judgement that the man is good is the conclusion. Whilst this way of understanding supervenience may lend support to Hare’s prescriptivism, which is a non-cognitivist theory, it does not support non-cognitivism *per se*. The cognitivist can accept this account and interpret the contents of the major premise and conclusion of the inference as being candidates for knowledge.

‘St. Francis was a good man’ and ‘*S* is the standard for judging whether a man is good’ certainly seem like candidates for knowledge.

Blackburn (1993a pp. 137-43) argues that only non-cognitivism can explain the status of the supervenience thesis as a conceptual truth. Someone who calls something, *A*, good when it has a set of natural properties, *N*, is not competent with the concept ‘good’ if they fail to call something else, *B*, good which shares *N* whilst failing to point out some relevant difference between *A* and *B*. This principle is a conceptual truth. How do we gain knowledge of it? Since it is a conceptual truth, it cannot be via inductive generalisation, from observing moral properties in the world and discovering that their instantiation maps in particular patterned ways onto natural properties. Nor is this principle discovered by uncovering analytic connections between moral concepts and natural concepts for, as the modified OQA shows, there are no such connections to uncover.

Blackburn suggests that non-cognitivists have an easy explanation of supervenience. According to the non-cognitivists, the purpose of moral judgements is to guide agents so that they will desire to bring about states of affairs and to perform actions with particular natural properties. Success in achieving these purposes is only possible if one’s use of moral language is consistent. Saying on one occasion that lying is wrong but that on another it is permissible is, on the crudest form of non-cognitivist analysis, like saying ‘do lie, don’t lie’. This is self-defeating unless some relevant difference can be pointed out between the two cases: ‘do lie when captured by the enemy, otherwise don’t lie’. This non-cognitivist analysis can show why the supervenience thesis is a conceptual truth: anyone who denies the supervenience thesis does not understand the purpose of moral judgements; anyone who contravenes the supervenience thesis defeats that purpose.

The argument from supervenience is interesting, but it only provides an advantage for non-cognitivism over some forms of cognitivism. Importantly for my purposes, it does not provide any advantage over cognitivist accounts based on the practical reason approach. According to the practical reason approach, judgements about what one is morally required to do are judgements about what one has most moral reason to do. This is a conceptual truth: what we are morally required to do is what overall the moral considerations present in the case at hand require. Moral reasons are the morally relevant considerations – facts such as ‘that he was in pain’, ‘that I had promised to’, ‘that he refused medical treatment’. Since it is a conceptual truth that what

is morally required is based on how these considerations weigh up, it is a conceptual truth that the only thing that can render two differing judgements valid is a difference in the considerations on which they are based<sup>12</sup>. Therefore, the supervenience thesis is unproblematic for cognitivists who adopt a practical reason approach.

Non-cognitivists might argue against cognitivism by claiming that the purpose of asserting that actions of a particular kind are morally required is to persuade people to perform actions of that kind. They might say this is the only plausible explanation of the pervasive use of moral language, its application to action, and its use in rhetorical and motivational contexts. Non-cognitivists might claim that cognitivists have no explanation of these facts: if the purpose of moral language is to state facts or to describe then why is it so often used persuasively, to influence conduct? Why is it that saying ‘it wouldn’t be right’ can convince someone not to perform that action after all?

Despite its initial plausibility, this line of argument is a non-starter. There are at least four serious problems with it. The first is that it cannot account for moral deliberation and for reflexive moral judgements. On this account, what does one wonder when one asks ‘I wonder whether it would be right for me to  $\Phi$ ?’ And why does one  $\Phi$  when one has decided that overall one ought to  $\Phi$ ? Why is it that when one deliberates about what one ought to do that one takes the process seriously and is concerned with trying to get the answer right? The second problem is that if moral language is used merely in attempts to influence conduct it becomes very difficult to see why it would ever be effective. People would begin very quickly to see that use of moral language was empty manipulative rhetoric and would strip it of the privileged status that it currently enjoys. Thirdly, it would be strange if e.g. my saying ‘you are morally required to  $\Phi$ ’ was just a way of expressing that I want you to  $\Phi$ . If that is what I intend to express I can do so directly by saying ‘I want you to  $\Phi$ ’. Fourthly, noting that moral language is often used persuasively does nothing to undermine cognitivism unless it is supplemented with the dubious assumption that utterances used persuasively are necessarily incapable of being truth-apt or constituting knowledge. Suppose that you are going on a journey across extremely rugged terrain and I insist that you ought to buy a Jeep rather than a Mini. I may both be trying to persuade you to buy the Jeep and telling

---

<sup>12</sup> Assuming that there is a single right way of moving from a set of moral reasons to an overall moral judgement (or that any right way of doing so produces the same results). This is part of the cognitivist claim and is supported, rather than undermined, by the supervenience thesis.



you something that I know (or that I am justified in believing): that overall there is more reason for you to buy the Jeep than to buy the Mini.

## 1.6 Conclusions

Two arguments for non-cognitivism have been examined: the open-question argument and the argument from moral disagreement. It was concluded that cognitivists who adopt a practical reason approach are largely immune to these objections. However, moral disagreement sets a number of conditions of adequacy for metaethical theories. We must be able to allow for cases of genuine moral disagreement, provide a robust moral epistemology and be able to explain the prevalence of moral disagreement. We saw that the force of the open-question argument comes from the practicality of morality and that when the practicality of morality is formulated as ‘moral internalism’ it appears to conflict with cognitivism, at least if we accept the Humean theory of motivation. This constitutes a barrier to the central claim I set out at the beginning of the chapter, seeming to drive a wedge between the theoretical and practical objectivity of morality. I begin to tackle this problem in the next chapter by examining whether we should accept moral internalism.

## Chapter 2. Moral Judgement Internalism

### 2.0 Introduction

In chapter one a number of challenges to cognitivism were discussed and found inconclusive. I proposed that by adopting a practical reason approach we might establish a position that meets these challenges and reconciles the intellectual and practical aspects of morality that appear to be in tension. In this chapter the practicality of morality takes centre stage, with cognitivism taking the role of a background assumption. It is questioned whether moral judgements are necessarily motivating. Arguments for internalism are found inconclusive. I sketch an attractive and modest form of internalism based on the practical reason approach. Later (in chapter eight) I aim to establish this form of internalism via this approach. Here (in chapter two) it emerges that internalism is much harder to establish than internalists have believed. I defend my form of internalism against some preliminary objections. However, by the end of the chapter it emerges that this form of internalism is incompatible with the Humean theory of motivation.

### 2.1 The argument for internalism about moral judgements

Smith distinguishes between three internalist claims. Let us call the first “strong internalism” (Miller 2003 p. 221) and initially follow Smith in referring to the second as “the practicality requirement on moral judgement” (1994 p. 62) and the third as “rationalism” (*ibid.*):

Strong internalism:

If an agent judges that it is right for her to  $\Phi$  in circumstances C, then she is motivated<sup>1</sup> to  $\Phi$  in C. (Smith 1994 p. 61)

The practicality requirement on moral judgement:

---

<sup>1</sup> I use the term ‘motivated’ in the sense that being motivated to  $\Phi$  does not imply actually  $\Phi$ -ing or attempting to  $\Phi$ , or even attempting to  $\Phi$  when an appropriate opportunity arises. It involves having a motivation that could be outweighed by other motivational factors. However, I think that having a motivation to  $\Phi$  does entail that if one had no competing motivations then one would necessarily attempt to  $\Phi$  given a relevant opportunity.

If an agent judges that it is right for her to  $\Phi$  in circumstances C, then either she is motivated to  $\Phi$  in C or she is practically irrational (*ibid.*).

Rationalism:

If it is right for agents to  $\Phi$  in circumstances C, then there is a reason for those agents to  $\Phi$  in C. (p. 62)

Smith rejects strong internalism as “a manifestly implausible claim” (1994 p. 61) because it “commits us to denying that, for example, weakness of the will and the like may defeat an agent’s moral motivations while leaving her appreciation of moral reasons intact” (*ibid.*). Since the form of internalism Smith defends is the practicality requirement on moral judgement in this chapter I refer to this claim simply as internalism<sup>2</sup>.

Smith claims to have a decisive argument for internalism (1994 pp. 71-6). He writes:

[I]t is a striking fact about moral motivation that a *change in motivation* follows reliably in the wake of a *change in moral judgement*, at least in the good and strong-willed person. (1994 p. 71)

Since good and strong-willed agents are rational<sup>3</sup>, internalism entails that they become motivated to  $\Phi$  whenever they come to judge that it is right to  $\Phi$ . Externalists must also explain this ‘striking fact’. According to Smith they cannot. Externalists claim that practically rational agents can judge that  $\Phi$ -ing is right in circumstances C whilst remaining unmotivated to  $\Phi$  in C. Therefore, for externalists, the practical rationality of good and strong-willed agents is insufficient to explain their reliable conformance of

<sup>2</sup> Smith’s characterisation of internalism is most naturally read as a modal claim: that moral judgements *necessarily* motivate rational agents. Zangwill claims internalism should be characterised as a claim about essences rather than in modal terms. He writes: “It could be that every person necessarily has the distinct desires that externalists postulate to explain moral motivation. In that case, moral beliefs would be necessarily motivating without being essentially motivating” (2008a p. 52, c.f. Zangwill 2008b pp. 94-5). I agree with Zangwill on this point, and where possible I focus on whether moral judgements are *intrinsically* motivating rather than whether they are necessarily motivating.

<sup>3</sup> Moral judgements may fail to motivate owing to irrationalities such as weakness of will (Smith 1994 p. 61). Smith gives an open-ended list of the types of irrationality that may break the connection between moral judgement and motivation. This threatens to make Smith’s position trivial (Miller 2003 p. 221). An independent account of practical irrationality is required if internalism is to avoid becoming the trivial claim that agents are motivated by their moral judgements except when there is something about their psychology that makes it the case that they are not. This concedes everything the externalist was looking

motivation to moral judgement. To complete the explanation, externalists must mention an additional feature of those agents. According to Smith, the only option for externalists is to posit a general desire that such agents have to do whatever it is that they judge to be right. As Smith puts it, “the only motivational content capable of playing this role... is a motivation to do the right thing, where this is now read *de dicto*” (1994 p. 74). Dreier writes:

Suppose I kept a list of foods, changing my list from time to time and giving you hints about what was on the list... You might have a peculiar desire *de dicto*, whose content was: eat whatever is on the list... If you saw me writing ‘avocado’, you would want to eat some avocados, even though at the moment you can’t stand them (2000 p. 623)

Externalism makes the moral motivation of virtuous agents analogous to the motivation to eat whatever is on the list: in coming to judge an action is right they come to be motivated to perform it *just because* it has the property of rightness. This makes the moral motivation of virtuous agents derivative and makes moral motivation seem like a peculiar fetish. This goes strongly against our intuitions that good people “care non-derivatively about honesty, the weal and woe of their children and friends, the well-being of their fellows, people getting what they deserve, justice, equality and the like” (Smith 1994 p. 75). Smith draws a comparison with an argument made by Williams (1976) against impartialist conceptions of ethics. Williams discusses a man forced to choose between saving his wife and saving a stranger. Whilst it is permissible for the man to save his wife, it is inappropriate for thoughts about the permissibility of saving her to play any role in his motivation. If they do, he has ‘one thought too many’ since his wife rightly hopes her husband acts out of direct love and concern for her without the need to reflect on what would be permissible or most morally praiseworthy for him to do. His concern for her should be direct and non-derivative.

Smith argues that analogous considerations apply to moral motivation. An agent who derives their desire to help a drowning child from (a) the judgement that ‘in situations like this it is right to help drowning children’ and (b) the desire ‘to do what is right’ has ‘one thought too many’. Smith says, “it is constitutive of being a morally good person that you have direct concern for what you think is right” (1994 p. 76). One should save the drowning child because one cares for the child’s well being, because

---

for, because that something might well be the absence of a contingent desire. This motivational lack

one cares that the child survives, not because it is an action that would appear on a list of all the morally required actions.

## 2.2 Miller's objection to Smith's argument

According to Miller, Smith's internalism suffers from the very same shortcoming of which he accuses externalism:

According to Smith's internalist, if George judges that honesty is right, it follows, as a matter of conceptual necessity, that George will be motivated to be honest.... [O]n this account George's motivation to be honest is still derived: not, to be sure, from his belief that it would be right to be honest in conjunction with an over-arching non-derivative desire to do the right thing, but from the belief that it would be right to be honest *itself*. (2003 p. 225)

In response to this worry, Smith altered his argument against externalism from the complaint that their account characterised the virtuous agent's motivations as *derivative* to the claim that it characterised such motivations as instrumental (1996a p. 182). He therefore accuses externalism of characterising the virtuous person's consistent action in accordance with their moral judgements as a *mere means* to fulfil their own (contingent) desire to be moral. If this is true, it represents a genuine worry for externalism. However, both internalism and externalism now appear problematic, since both mischaracterise virtuous agents as having only derivative concern for things that have intrinsic moral value.

Internalists might maintain that their theory is less unsatisfactory, however. Whilst internalism makes the motivations of virtuous agents derivative from their moral judgements, virtuous agents are motivated by their moral judgements alone, and their actions are not means to fulfilling other desires that they have, they are non-instrumental. Miller thinks this reply is insufficient, because internalism still leaves the virtuous agent with 'one thought too many'. Returning to the analogy with Williams' example, Miller writes:

[T]he woman's complaint is that there is some aspect of the husband's psychology, in addition to the belief that the woman drowning is his wife, to which his motivation to jump in and save her is sensitive, where, crucially, it doesn't matter whether or not the other aspect of the husband's psychology is a desire or whatever. (2003 p. 227)

---

cannot be called 'irrational' without some justification being provided for this claim.

Applying this to Smith's internalism, Miller concludes that on the internalist account:

[T]he morally perfect person's desire to be honest... is still sensitive to some other aspect of the morally perfect person's psychology, namely the belief that it is right to be honest. As in the case of the husband with 'one thought too many', this is enough to merit the charge that the psychology of the morally perfect person has been misdescribed: in the morally perfect person, the motivation to be honest would be sensitive to no such aspect. (2003 p. 227)

Miller's conclusions do not seem quite right. It doesn't seem true that the drowning wife's complaint is that the husband's motivation is sensitive to some aspect of his psychology other than his belief that the drowning woman is his wife. This is highlighted by the fact that, for various reasons, some marital relationships do not fit the picture presupposed in Williams' example. As Wolf (2009) says, it may be that the man's wife has been plotting to kill him, that their marriage is a sham, that he never wanted to marry her, and that he cannot stand her presence. The wife's hope that her husband is motivated by the thought 'that is my wife' is based upon, and realistic only if, their marriage is one of love and commitment. The woman hopes that no other explicit thought than 'that is my wife' is necessary to motivate her husband. However, she does not hope that her husband's motivation is sensitive to no other aspect of her husband's psychology. The fact that he loves her (and many other facts about their relationship and his feelings and commitment towards her) makes an enormous difference to his psychology, a difference that she hopes plays an explanatory role in why he acts as he does<sup>4</sup>. Therefore Miller's conclusion that "the woman's complaint is that there is some aspect of the husband's psychology, in addition to the belief that the woman drowning is his wife, to which his motivation to jump in and save her is sensitive" (2003 p. 227) is false. Miller cannot use this example to drive his objection to the argument for internalism.

Has Miller, nonetheless, identified a difficulty for Smith's position? He claims that Smith's internalism is problematic because the morally perfect person's desire to be honest, for example, should not be dependent on their belief that it is right to be honest. If someone is morally perfect then they will (want to be honest and) be honest just because they care about honesty or see the value of honesty. If their motivation to be

---

<sup>4</sup> Note also that when the husband is motivated by the thought 'that woman is my wife' his motivation is not an independent aspect of his psychology; it is ontologically dependent on the belief in question.

honest is dependent upon anything at all, it might be argued, it should be dependent upon (an understanding of) the various reasons for being honest in different situations. The motivation should *not* be simply dependent upon the belief that it is right to be honest: this is a fetish, not a virtue. In advocating internalism based on a practical reason approach, I largely agree with these claims (as I explain in § 2.6.)

So far we have seen that internalism apparently characterises good agents as having a concern for things that have intrinsic moral value only because of their moral beliefs. Moral goodness appears to be compatible with lacking direct, non-derivative concern. Externalism appears to be in a worse state, making moral virtue into a fetish whereby good moral agents have a peculiar desire to tick off items on a moral shopping list simply because they carry the ‘morally right’ brand label. Moreover, they are only concerned with doing what is morally right because it helps them to satisfy this peculiar desire that they have. Internalism appears unsatisfactory, but externalism seems even worse. It makes morality into a strange collective fetish<sup>5</sup>. However, externalism can overcome this problem by uncovering and rejecting a mistaken assumption. I turn to this in the next section.

## 2.3 Virtue and the motive of duty

There is an assumption present in Foot (1972) common to Smith’s criticism of Foot’s externalism and Miller’s criticism of Smith. We can distinguish, as Kant did, between action from duty and action from sympathy or ‘compassion’:

To be beneficent where one can is a duty; and besides this, there are many persons who are so sympathetically constituted that, without any further motive of vanity or self-interest, they find an inner pleasure in spreading joy around them and can rejoice in the satisfaction of others as their own work. But I maintain that in such a case an action of this kind, however dutiful and amiable it may be, has nevertheless no true moral worth (1785 § 398).

Acting from duty (moral motivation) can be differentiated from being moved by sympathy (altruistic motivation). Foot notes that Kant “contrasts acting out of respect for the moral law with acting from an ulterior motive” (1972 p. 313) and claims that “Kant... was a psychological hedonist in respect of all actions except those done for the

---

<sup>5</sup> However, it might actually be thought *admirable* for someone to have a general motivation to be moral. After all, this seems similar to Kant’s notion of a good will, which he takes to be the only thing “which can be regarded as good without qualification” (1785 §393). This idea emerges in §2.3

sake of the moral law (*ibid.*). Whether this was Kant's view or not, it is certainly related to a mistaken assumption underlying the internalism debate: the assumption that moral motivation must come either from moral judgements themselves or from some non-moral source. As Zangwill points out, there is another alternative: moral motivation might come from desires with specifically moral content. Such desires would link up with moral judgements to provide motivations to perform the actions about which those judgements were made. Zangwill calls this form of externalism, which he defends, "Moral Content Externalism" (2003 p. 145).

According to this Moral Content Externalism, moral judgements are beliefs that motivate us when combined with desires with moral content. A desire with moral content is such that "its content represents a mind-independent moral state of affairs, where a moral state of affairs is the instantiation of a moral property by something" (Zangwill 2003 p. 144). So, for example, a desire 'to be honest' might combine with a judgement that 'telling her that I lost the money was the honest thing to do', to produce a motivation 'to tell her that I lost the money'. This would explain why the motivations of good and strong-willed agents track their moral judgements: such agents have the desires with the right moral contents to connect with moral judgements and produce motivation.

A virtuous person could "have a battery of non-derivative desires, each of whose content is quite broad" (Shafer-Landau 1998 p. 356). For example "a quite general desire to support just political institutions, or more egalitarian political parties" (*ibid.*) might explain why an agent's motivations track one subset of their judgements about what they morally ought to do. Similarly, "one might have fundamental, non-instrumental desires to see one's family flourish, to see justice done, to promote the welfare of the worst off in society, to work towards gender equality" (*ibid.*). That good and strong-willed agents have such desires is not implausible. After all, it could plausibly be suggested that agents who are good and strong-willed are those who have acquired such a set of desires through successful processes of moral education. Therefore externalists do not need to posit a single general desire to do what is right in order to explain the changes in motivation of good and strong willed persons.

The externalist strategy that we have seen is that of defusing the internalist challenge by opting for the first horn of a dilemma posed – but not emphasised – by Smith. The dilemma is to choose between two possible explanations of what happens



when the good and strong-willed agent's motivation changes in the wake of a change in moral judgement:

(a) [H]er motivation stems from the fortuitous linking up of her new moral belief and some pre-existing desire, or (b) her motivation is generated derivatively from an overarching desire to do what is right, understood *de dicto* (Shafer-Landau 1998 p. 355)

Smith dismisses the first option because it seems implausible and *ad hoc* that, for each change of moral judgement, the good and strong-willed person just happens to have some desire that would be satisfied by acting in accordance with that new judgement. Yet this is not implausible at all; moral education may provide us with the relevant moral desires.

Externalists may yet be pushed onto the second horn of the dilemma. However, when we see why, we will see that in fact the second horn is blunted as well, and that the best externalist strategy is to embrace both possibilities as complimentary answers to the question of how agents come to be motivated by their moral judgements. The first option – that motivation stems from pre-existing desires – cannot explain “how new moral beliefs may overthrow standing fundamental desires, or supply motivation in the absence of any relevant antecedent desire. Here externalists must invoke the motive of duty to explain the acquisition of new desires” (Shafer-Landau 1998 p. 356). However, it is unproblematic that in some circumstances – such as changes in fundamental values – agents should act upon the motive to do what is right, understood *de dicto*. Most of the time morally decent agents do not need to act from a motive of duty because they have appropriate moral habits and dispositions. However, during reflection on their habits, dispositions, motives, commitments and principles, it may be perfectly morally acceptable, or even commendable, that their reflection refers explicitly to rightness itself. There is nothing wrong, it seems, with an agent deciding that they ought to adopt a moral principle or cultivate a certain moral disposition because they believe it right that they do so, where this is understood *de dicto*. Shafer-Landau characterises good agents as “defined in part as people who use the standard of rightness, whatever it is, as a kind of limiting condition on the formation of their motives” (1998 p. 357). He envisages the conscious employment of explicit standards of rightness as a restraint on the development of their moral character and commitments, such that “were a concern to appear to them as immoral, then they would abandon or modify it (or reassess their

standard of rightness)” (*ibid.*). Therefore, being guided by motives of duty has its place in the moral life.

Externalism has not quite escaped the objection yet. Zangwill says that Smith’s objection “is that externalism implies that the ‘virtuous agent’ is ultimately motivated by the fact that he possesses a moral property rather than by the natural features of the situation in which he acts” (Zangwill 2003 p. 148)<sup>6</sup>. However, our *attributions of virtue* to an agent do not necessarily need to be based on facts about how moral judgements motivate that agent. Zangwill suggests that moral desires are always necessary for action on moral beliefs, but that non-moral motivations (altruistic motivations) are what make us morally virtuous (2003 p. 149). Therefore, externalism escapes Smith’s objection. Smith seems to imagine that the only plausible account of moral agency must be one in which the moral agent always acts out of altruism (direct concern, sympathy) but that altruism is somehow identical with or guaranteed by moral judgement. This may be an appealing picture, but it is not the only possibility, and it is not established by the arguments that Smith gives. Externalists are able not only to blunt both horns of the dilemma, but to show that they do not really have to choose between the two options: they can, and should, have both.

## 2.4 Thoughts, desires, and dispositions

In addition to the arguments already presented, there is another powerful reason for thinking that Smith’s argument against externalism cannot succeed. Smith argues that the externalist characterisation attributes to the virtuous person ‘one thought too many’. However, it seems that, given his defence of the Humean theory of motivation, this accusation cannot be sustained. Smith (1994 pp. 92-129) argues that we must accept the Humean theory and that in order to do so we must accept a dispositional conception of desire. He argues that, in any case, we ought to accept a dispositional conception of desire because it is the only way to make sense of the epistemology of desire (pp. 105-7), the phenomenology of desire (pp. 108-9) and the idea that desires have propositional content (pp. 107-8). Smith claims that:

[W]e should think of desiring to  $\Phi$  as having a certain set of dispositions, the disposition to  $\Psi$  in conditions  $C$ , the disposition to  $\chi$  in conditions  $C'$ , and so on, where, in order for conditions  $C$  and  $C'$  to obtain, the subject must have *inter*

---

<sup>6</sup> Cf. Smith 1994 (pp. 72–76); Smith 1997 (pp. 112–117)

*alia*, certain other desires, and also certain means-ends beliefs, beliefs concerning  $\Phi$ -ing by  $\Psi$ -ing,  $\Phi$ -ing by  $\chi$ -ing and so on. (Smith 1994 p. 113)

Therefore, according to Smith, any particular desire may or may not be available to introspection and may or may not have phenomenological content. However, all desires dispose agents to action. Desires are *essentially* dispositional states. They do not need to play any role in conscious thought or reasoning processes; very often they may just be dispositions to act in certain ways under certain conditions. Therefore, the general desire to do what is right that externalists are forced to posit to explain the motivations of the virtuous agent may well be a desire that the virtuous agent is never consciously aware of. It may not play any part in their reasoning at all. However, the idea that externalism made virtue a moral fetish was dependent on the thought that externalists are forced to characterise the virtuous agent as reasoning in the following manner:

I want to do whatever is right; I have a general desire to do what is right.

It is right to save this child's life; saving this child is a way to do what is right.

So, I'll save this child – that way, I'll satisfy my desire to do what is right.

However, if we accept the dispositional conception of desire, as Smith does, when the virtuous agent is motivated by a belief of the ' $\Phi$ -ing is right' type, there might be nothing more to be said about their thoughts<sup>7</sup>. Rather, they might just immediately find themselves disposed towards performing acts of type  $\Phi$ . Hence, the charge that externalism leaves the virtuous agent with 'one thought too many' seems unfounded.

Copp (1997 p. 50) suggests an externalist possibility along these lines (although he does not mention its relation to Smith's commitment to the dispositional conception of desire). Copp suggests that externalists can explain why good and strong-willed agents desire to do what they judge right by attributing to those agents a disposition:

[T]he disposition is not the *de dicto* desire *to-do-whatever-is-right*. For it is not a desire from which a person would *derive* any further desires to act. If a person has the disposition I have in mind and comes to believe that she would be right

---

<sup>7</sup> Shafer-Landau writes: "What seemed especially repugnant about the externalist commitment was the thought that a good agent's concerns for family, for honesty and for justice must be derivative from concern with moral rightness. But this is offensive only if derivative concerns must be self-consciously generated from one's direct concern(s)" (1998 p. 358). However, Shafer-Landau's does not recognise the possibility that the motivational force of moral judgements might come from a single desire without the agent having 'one thought too many'.

to do  $\Phi$ , she then desires straightaway to do  $\Phi$  without deliberation or inference, conscious or unconscious (*ibid.*).

Since this explanation is available whether desires are dispositions or not, Smith's argument could not succeed even if he abandoned the dispositional conception of desire. Good and strong-willed agents may, in Dreier's words, be "morally suggestible" (2000 p. 624); they may just find that, upon coming to believe something is right, they are motivated to do it. However, the production of this motivation might be not due to their moral judgements alone, but to those judgements in combination with a disposition.

## 2.5 Moral rationalism

The considerations presented so far have all proved indecisive. However, Smith has another argument for internalism that is based on moral rationalism. Smith claims that "our moral judgements are expressions of our beliefs about what we have normative reason to do" (1994 p. 185). On this view, moral judgements are judgements about what one has most moral reason to do. Insofar as one is practically rational one is motivated to act by the reasons for action that one takes there to be<sup>8</sup>. If so, then when one judges that there are moral reasons to  $\Phi$  one is motivated to  $\Phi$ . On this account, one's motivation to  $\Phi$  tends to come into existence with one's belief that it is right to  $\Phi$  and go out of existence when that belief goes out of existence, but this does not mean that one is motivated by the thought 'it is right to  $\Phi$ '. Rather, in such cases, if one is rational, one's desire to  $\Phi$ , one's judgement that it is right to  $\Phi$ , and one's judgement that there are moral reasons to  $\Phi$  are all based on one and the same thing: the reasons present in the situation at hand.

This internalist thought avoids the problems of (a) making moral virtue into a fetish (b) making moral motivations derivative (c) making moral motivations instrumental in nature and (d) attributing to the virtuous agent 'one thought too many'. If this is right then internalism, like externalism, can provide a plausible account of how changes in moral judgement lead to changes in motivation in good and strong-willed agents. This account of moral judgement might even be taken to provide a new argument for moral internalism, so long as there is some good argument for the moral rationalism that it takes as its premise. We can set out this argument as follows:

---

<sup>8</sup> Smith does not argue for this since he takes it to be a 'platitude' about practical rationality. Whether this claim is a 'platitude' or not, I think it requires some explanation and support. I provide this in § 2.6.

1) Moral rationalism:

Our moral judgements are expressions of our beliefs about what we have normative reason to do. (Smith 1994 p. 185)

2) Practical rationality:

Insofar as one is practically rational, one is motivated by what one believes there to be (normative) reason to do.

∴ Moral internalism:

Insofar as one is practically rational, one is motivated by one's moral judgements.

This argument is valid<sup>9</sup> but moral rationalism is a controversial claim. To get this argument for internalism to work we need a good argument for moral rationalism. Smith claims to give just such an argument which begins from the idea that “requirements of rationality or reason are categorical imperatives” (1994 p. 85). The argument is as follows:

1. Requirements of rationality or reason are categorical imperatives
2. We expect that rational agents will do what they are morally required to do
3. Therefore, since it is their being rational alone that grounds this expectation, it must be that their being rational alone is enough to guarantee that they do what they are morally required to do.
4. Therefore, it must be the case that requirements of morality are requirements of rationality.
5. Since requirements of rationality are categorical imperatives, requirements of morality must be categorical requirements of rationality.

However, it seems that, as it stands, Smith's argument is utterly unconvincing. Whether or not the argument succeeds depends on the truth of the claim that links the second and

---

<sup>9</sup> Copp (1997) argues that given Smith's account of what it is for something to be a normative reason this argument isn't actually valid. I ignore this complication here and hold the argument to be valid. This does

third premises: that the fact that an agent, *A*, is rational gives us sufficient grounds to expect that they will do what they are morally required. Crucially, this depends on whether or not being rational should be characterised as necessarily involving being moral. This in turn depends upon whether moral requirements are rational requirements. However, this was what the argument was supposed to establish. Therefore, as it stands the argument is viciously circular.

Smith attempts to buttress his basic argument with two additional considerations. The first is that “we certainly expect rational agents to do what they *judge* themselves to be morally required to do: that is, we certainly believe not just that they should, but that they will, other things being equal” (1994 p. 86). This is the ‘practicality requirement’ (see § 2.1), and it is a statement of internalism. Since we are trying to see whether we can support internalism via moral rationalism, to appeal to internalism to support rationalism would be futile. It is therefore useless for our purposes.

Smith’s second buttressing consideration is that:

- (a) “we approve and disapprove of what people do when moral matters are at stake” (1994 p. 89) and
- (b) “approval and disapproval are only ever in place when there exist *grounds* for legitimate expectations about how someone will behave” (*ibid.*)

It then follows, therefore, that:

- (c) when moral matters are at stake, there exist grounds for legitimate expectations about how someone will behave

From this, Smith concludes:

- (d) “What grounds the legitimacy of our expectation is the mere fact that people are rational agents. Being rational suffices to ground the expectation that people will do what they are morally required to do.” (1994 p. 90)

---

not undermine my own position because do not rely on this argument and I do not accept Smith’s account of normative reasons.

Even if this argument were valid, its conclusion would not be established, because the second premise (b) is false. It is quite possible for racists to disapprove of measures designed to stamp out institutionalised racism, or for holocaust deniers to disapprove of the history syllabus in British schools. It does not thereby follow that there exist grounds for legitimate expectations that measures designed to stamp out institutionalised racism not be taken and that all mentions of the holocaust be removed from history lessons in British schools.

The premise (b) can be held true if it is taken to be a premise about *legitimate* approval and disapproval. Then, however, for there to be any hope of the argument being valid the first premise (a) needs to be modified to state that

(a\*) We *legitimately* approve and disapprove of what people do when moral matters are at stake.

Roughly speaking, it is only legitimate to disapprove of someone's action when overall there was most reason for them not to act as they did or they acted on the basis of considerations that were not good reasons for action. This is because disapproving of something involves thinking that it should have been otherwise or it would have been better if it were (or could have been) otherwise. This in turn involves thinking that there were reasons for things to be made different or done differently. In the case of the disapproval of *actions* this involves thinking that reasons for action were ignored. Therefore, moral disapproval is legitimate when moral reasons for action go unheeded. Therefore, the claim that we legitimately disapprove of people when moral matters are at stake presupposes that there are moral reasons. It therefore presupposes that there are reasons for us to conform to moral requirements, and hence it presupposes a form of moral rationalism. As an argument for moral rationalism it makes no progress.

Another way of putting this is to point out that the legitimacy of moral approval and disapproval would only have to be grounded in the fact that people are rational agents if moral approval and disapproval are rationally required responses to instances of rational success and failure respectively. This would only be the case if moral requirements are rational requirements. However, we do not have any reason yet to think that they are – for this was exactly the point that Smith was looking to establish<sup>10</sup>.

---

<sup>10</sup> Miller (2003 pp. 231-2) makes a similar point about Smith's argument.

Therefore, Smith has no good argument for internalism, whether via moral rationalism or independently of moral rationalism.

## 2.6 Internalism and the practical reason approach

Smith's attempt to establish internalism by appealing to normative reasons seems to be along the right lines. As Wallace says, "[w]hat is attractive about internalism regarding moral judgment... is the thought that moral considerations at least purport to have normative significance" (2006b p. 186). We should think of normativity in terms of reasons but not necessarily in terms of rationality (see § 1.3). Roughly speaking, being rational is ordinarily thought of as a matter of adopting reliable, valid and effective patterns of reasoning and of being responsive to the reasons for belief and action that one recognises. We are *rationally required* to do whatever constitutes being rational in this sense. It is not clear whether or not morality is partially constitutive of this or not. However, moral requirements do not need to be rational requirements in order for morality to be normative. For morality to be normative there need to be moral reasons. When we make moral judgements we make judgements about what the moral considerations, taken together, require. If there really are moral reasons then these judgements, when correct, report normative claims on our conduct. I propose, therefore, that instead of understanding internalism as Smith does, we should understand internalism according to the practical reason approach as a claim closely related to the practical objectivity of morality. We should see it, more specifically, as the claims that:

- (i.) moral requirements apply where, and only where, there are moral reasons
- (ii.) judgements about moral reasons can, and should, be *intrinsically* motivating<sup>11</sup>
- (iii.) overall moral judgements do not necessarily provide or make appropriate any motivation above and beyond the motivation provided by moral reasons

The first and second of these claims affirm that moral requirements are normative. The second claim also indicates how it is that moral requirements motivate if indeed they

---

<sup>11</sup> Mason (2008 p. 139) claims that saying that reasons have motivational implications assumes internalism about reasons. However, saying, as I do, that *recognising* reasons has motivational implications is a different claim and does not assume internalism about reasons. Internalism about reasons is the claim that a consideration is a reason for an agent only if it is suitably related to some motivation of the agent. I discuss this controversial claim in chapter seven and conclude that it is false.



are normative. If there are moral reasons then this second claim can be justified as follows. When one judges that a particular consideration is a reason for action this itself constitutes a rational commitment to action. Judging that  $r$  is a reason to  $\Phi$  rationally commits one to the judging that if there were no other considerations relevant to what one should do then overall one has most reason to  $\Phi$ . Judging that overall one has most reason to  $\Phi$  rationally commits one to  $\Phi$ -ing. Statements such as ‘I knew it would be best overall to catch the train, but I drove instead’ or ‘I knew that, overall, all the relevant considerations favoured taking out travel insurance, but I decided not to’ are paradigmatic examples of confessions of irrationality. In such cases, the question ‘why?’ is not a request for a justification of the decision, not a request for more reasons: if any more reasons could be given then the judgement was not, after all, an overall judgement. To ask why someone acted contrary to what they took themselves to have most reason to do is to seek the *cause* of their *irrationality*. When we are thinking rightly, the reasons we recognise motivate us simply because we recognise them *as* reasons. To recognise something *as* a reason for action is to take it to be relevant to what one does, and to do this is to consider it in one’s deliberations about what to do<sup>12</sup>. One can recognise a consideration that is a reason without recognising that it is a reason. Conversely, one can believe there is a reason without having any belief about what the reason is. Garrard and McNaughton write:

I may not only believe that you are in pain, but also believe that the fact that you are in pain gives me a reason to give you an analgesic. (1998 p. 54)

Your pain is the reason for me to give you a painkiller; I am motivated by my recognition of your pain and of the fact that it gives me a reason to give you a painkiller. If I recognise that you are in pain without realising that this gives me a reason to give you a painkiller then I may not be motivated to give you a painkiller. Recognising a feature that provides a reason can come apart from recognising *that* that feature provides a reason, as Garrard and McNaughton point out:

This seems to be what happens in the case of the person who has become depressed; she may retain her belief that, for example, if she doesn’t act now she

---

<sup>12</sup> A defeated reason should not be considered motivating, since the agent no longer considers it to have any normative force. However, a reason that is outweighed should still be considered motivating even though it does not explain any action of the agent. Outweighed reasons are still taken to have normative force and are also such that they *could* have led an agent to action if it were not that they had stronger reasons to do something else.

will lose this opportunity for which she has been working for two years, but she no longer believes that this gives her a reason to act. (*ibid.*)

If one is rational, when one deliberates, one's deliberations determine what one does. Since what determines what one does is what motivates one, when one takes a reason into account in deliberation one is motivated by it. Therefore, rational agents are *necessarily* motivated by their judgements about reasons. This includes moral reasons, if there are any.

On the form of internalism just sketched it is not obviously true that overall moral judgements are intrinsically motivating. An overall judgement that  $\Phi$ -ing is morally required certainly indicates that there are good reasons to  $\Phi$ , and these reasons to  $\Phi$ , if recognised, will be intrinsically motivating for rational agents. Nonetheless, there are two reasons to be hesitant about saying that the overall judgement is intrinsically motivating. Firstly, as Dancy points out, the fact that an action is right is not itself a reason<sup>13</sup> to do it:

It is the features that make the action good or right that are the reasons for doing it, and to say that it is good or right is merely to express a judgement about the way in which other considerations go to determine how we act. (2004 p. 16; cf. Garrard and McNaughton 1998 pp. 53-4)

This said, in deliberating about our moral duties we may<sup>14</sup> move from contributory moral reasons to an overall moral judgement and from this judgement to action. In some cases where moral considerations guide our actions it may be that the outcome of our deliberation about what our moral duty is determines what we do. In such cases it is appropriate to think of overall moral judgements or the act of overall moral judgement as being intrinsically motivating. Even if such judgements only<sup>15</sup> synthesise intrinsically motivating contributory reason judgements, in cases where the overall judgement leads to action the fact that we say that the contributory reasons motivated the agent is dependent on the act of overall judgement, since the decision to act was so dependent.

---

<sup>13</sup> That is, it is not an *extra* reason to do it. This may seem to beg the question against a Kantian view according to which there are some actions that we ought to perform because they are commanded by the moral law. I discuss how such views relate to the practical reason approach in chapter eight (§§ 8.8-8.9).

<sup>14</sup> Our deliberation does not necessarily need to go via an overall moral judgement. It could go via an overall judgement of what there is most reason to do or it could, perhaps, move to action without going via either of these kinds of overall judgement.

<sup>15</sup> Arguably they do not: the outcome of overall moral judgement is not achieved by merely adding together the weights of the contributory reasons favouring each alternative and determining which alternative is backed by the greatest weight of reasons (Dancy 2004).

Therefore, in such cases, overall moral judgements do appear to be intrinsically motivating.

There is a second reason, however, for hesitating to say that overall moral judgements are intrinsically motivating for rational agents. According to the view I will defend, moral considerations are important but not overriding. Therefore, even when one morally ought to  $\Phi$ , one might have most reason to do something else. In such cases the contributory moral reasons on which the overall moral judgement is based should still be considered motivating<sup>16</sup>. However, the judgement that the moral considerations taken together overall favour  $\Phi$ -ing is not itself a reason to  $\Phi$ . Therefore, in cases where this judgement does not lead to action due to the presence of weightier non-moral reasons, it cannot be said to be intrinsically motivating.

The resultant form of internalism is one according to which judgements about moral reasons are intrinsically motivating for rational agents and we *ought* to be motivated *in accordance with* our moral judgements. Overall moral judgements can be intrinsically motivating but they are not necessarily motivating, even for rational agents. This apparently strange<sup>17</sup> consequence results from differences in moral *understanding*. Some people do not think that there are any moral reasons and think moral requirements provide no reason for action. If my form of internalism is correct, it will show that they are mistaken. However, it will not necessarily show that they are irrational. It will show that they do not understand the nature of morality, they do not understand that morality is normative, reason-giving. Likewise, agents who do think that there are moral reasons may fail to fully understand morality or to fully appreciate the normative force of moral claims. They might also, without any irrationality, make mistakes about the moral reasons there are. I do not think that in these cases we should say that such agents do not really make moral judgements. Most of us most of the time fail to fully comprehend the moral reasons we have (often through no rational fault of our own.) Sometimes when we recognise moral reasons we are inappropriately dismissive of them, and sometimes this is due to the fact that we do not take the time to reflect on their significance. This is

---

<sup>16</sup> For reasons explained above this includes both those that were outweighed and any that, together with non-moral considerations, favoured the action that there was most reason to do.

<sup>17</sup> Dancy (1993 pp. 23-4) expresses a similar view about which Zangwill expresses puzzlement. Zangwill writes: "If water is *essentially* H<sub>2</sub>O, then water is *necessarily* H<sub>2</sub>O and *all* water is H<sub>2</sub>O, and similarly if moral beliefs are *essentially* motivating, then they are *necessarily* motivating and they are *all* motivating, whether or not they lead to action" (2008a fn. p. 54). However, some moral beliefs might be intrinsically motivating without it being essential to moral beliefs that they are motivating. I claim that when a moral belief is held but not understood it may fail to motivate, but that moral understanding *is* essentially motivating.

one reason why we do immoral things that we regret: afterwards we have plenty of time to reflect and to recognise the full force of the moral considerations that in the heat of the moment we failed to reflect on and impatiently dismissed. Without moral understanding, moral judgements may fail to motivate, but when moral understanding is present moral judgements involve the recognition of moral reasons *as* reasons. For rational agents, this is motivating. Moral understanding together with rationality ensures that moral judgements are accompanied by motivation<sup>18</sup>.

## 2.7 Volitional impossibility

I have sketched a form of internalism based on the practical reason approach, claiming that if we can demonstrate that there are moral reasons then this account will be vindicated. However, showing there are moral reasons is no easy task, so, if anything, rather than providing substantial support for internalism, I have shown how difficult internalism will be to establish. Even so, there are three immediate objections to what I have said so far. The first arises from the phenomenon of volitional impossibility. The second arises from the claim that normative considerations do not necessarily have motivational implications. The third arises from the possibility of variable motivation, amorality and cynicism about morality.

Miller (2008) thinks that cases of volitional impossibility undermine internalism. Frankfurt gives the following example:

Consider a mother who reaches the conclusion, after conscientious deliberation, that it would be best for her to give up her child for adoption, and suppose that she decides to do so. When the moment arrives for actually giving up the child, however, she may find that she cannot go through with it – not because she has reconsidered the matter and changed her mind but because she simply cannot bring herself to give her child away. (1993 p. 111)

We might change the example to imagine the mother originally judging that it was both morally best and best all things considered to give her child up, and later changing her mind about the latter judgement but not the former (Miller 2008 pp. 238-9). Through a “newfound depth of care for her daughter” (p. 240) she may display, and fully endorse,

---

<sup>18</sup> Arguably virtue is also required. However, in the context of stating the internalist position it does not need to be explicitly mentioned. If internalism is correct then moral understanding and rationality are partially constitutive of virtue; they are the parts that suffice for moral motivation to accord appropriately with moral judgement. Internalism does not purport to give a full account of what it is to be a morally ideal (virtuous) agent.

an aversion to giving up the child. She continues to judge it morally best to give up the child, but she displays no irrationality and utterly no motivation to go through with it.

Reflection on this example damages many forms of internalism, but not all. An internalism that is happy to admit that moral considerations are not overriding and can be defeated is not threatened by the possibility of such instances of volitional impossibility. Since morality is directly concerned with the well being, value and rights of persons, it is highly plausible that moral requirements might be defeated in many cases where people find themselves incapable of bringing themselves to perform an action. A mother's love for her daughter seems like exactly the kind of consideration that can defeat moral reasons. This indicates not that there are no moral reasons but that a moral consideration that appears to have normative weight might turn out not to, and that a feature that is a moral reason in one situation might turn out to be no reason at all if circumstances change<sup>19</sup>.

## 2.8 The motivational implications of normative considerations

Svavarstóddir says that even if morality is inescapable, its inescapability is normative rather than motivational, and that it is possible to be rational and acknowledge the normative significance of something without being motivated. She writes:

I acknowledge that it is incumbent on me to act in accord with the local traffic laws and, hence, to take my sincere judgements about those laws as a guide to conduct. Yet, I am not inclined to take them wholesale as *my* guide to action. For example, I have my own policy on jay-walking: jay-walk with caution except in front of young children. (2006 p. 174)

What this shows is that one can be rational whilst being only motivated to act in accordance with a particular requirement insofar as one takes it to provide reasons for action. Svavarstóddir apparently thinks that the requirement not to jay-walk derives its reason-giving force (roughly speaking) from two facts. Firstly, incautious jay-walking is dangerous; secondly young children are not yet able to judge the risks in jay-walking so should be encouraged not to do it at all. Therefore, Svavarstóddir seems to think that a requirement is normative just insofar as it identifies reasons. This seems right, and it

---

<sup>19</sup> These claims express holism about reasons. Holism about reasons is now widely accepted. Whilst holism about reasons has been claimed to support moral particularism (Dancy 2004) it is unclear whether it does (McKeever and Ridge 2006). Holism is generally accepted even by those who reject particularism (see Garrard and McNaughton 1998 p. 50).

indicates that whilst Svavarstóddir's observations undermine some forms of internalism, they do not undermine an internalism based on the practical reason approach.

Svavarstóddir questions whether saying that the fact that jay-walking is illegal is a consideration that should be taken into account when:

- (1) deciding whether to jay-walk on a particular occasion, or
- (2) deciding what personal policy to adopt on jay-walking, or only
- (3) as relevant to arguments about whether one should jay-walk

It is unclear why this is a problem: there can be considerations that are reasons for action on a particular occasion, others that are reasons for adopting particular policies, and others that are both. Any of these kinds can be offered in arguments about whether one should jay-walk. Svavarstóddir seems to imagine that individuals differ on how they take reasons for action in account, and that there is no ground for suggesting that they shouldn't. She thinks that given her policy of jay-walking on some occasions, which we have agreed is rational, she is only motivated by the reason in the second and third contexts. However, I think Svavarstóddir is mistaken. What she says suggests that she *never* takes the fact that jay-walking is illegal to be a reason for action. She thinks that the reasons for not jay-walking are provided by the danger of incautious jay-walking and the inability of young children to judge when jay-walking is safe. Such reasons would apply even if jay-walking were legal (if they are reasons at all then they do actually apply in countries where jay-walking is legal, such as the UK). They also apply in all three contexts: for assessing particular acts, deciding upon policies, and arguing about whether one should jay-walk.

Contra Svavarstóddir, recognising a reason for action seems necessarily to involve taking it as *my guide* to action. Garrard and McNaughton write:

There is clearly a difference between saying, perhaps sincerely, that one believes one has reason to A and actually believing that one has reason to A (for people can be mistaken about their beliefs). In what does this difference consist? It is surely tempting to say that one who does really believe that she has reason to A is inclined to A, cares about A-ing etc. For, given that this is a belief about a practical reason, in what other way could the presence of such a belief manifest itself? (1998 p. 53)

To put the point slightly differently: a reason for action is a consideration that is to be taken into account when deciding what to do. Whether someone takes *r* into account in deciding what to do when in situations of type *S* is a key indicator of whether or not they genuinely believe *r* to be a reason for action in situations of type *S*<sup>20</sup>. If thinking that one has a reason for action bears no relation to one's reasoning about how to act then it is hard to see in what sense one thinks it is a reason for action at all<sup>21</sup>.

## 2.9 Variable motivation

Mason says, “the externalist ... should say that it is possible to accept that she has reasons (including moral reasons), and yet not be motivated” (2008 p. 139) *or*, if they admit that this is not possible, they “could deny that moral ought statements are statements of reasons” (*ibid.*). We have just rejected the first possibility, at least as a thesis concerning rational agents. However, externalists still have the second possibility. Amoralists may be possible. Before coming to amoralists, I want to consider an objection from Zangwill relating to variable motivation.

Zangwill says that to capture the spirit of their position, internalists should add that moral judgements bring with them the strength of motivation appropriate to the confidence with which the judgement is held (2008a p. 53; 2008b p. 95). Otherwise it would be possible for a tentatively held judgement to be more motivating than an equivalent judgement held confidently. Is this right? If your pain is a reason for me to give you a painkiller, it is just as much of a reason whether my belief about your pain is tentative or firm. Nonetheless, from the first person perspective things are more complicated. Plausibly, there are norms in play relating the confidence of our beliefs about reasons to the weight that we should assign to those reasons in deliberation. Moreover, we should add that since reasons also differ in the strength, internalists should maintain that agents are not just required to be morally motivated, they are required to be morally motivated to the appropriate extent. Moral reasons should receive the appropriate weightings in deliberation<sup>22</sup>.

---

<sup>20</sup> The relation is normative: irrational people can believe they have a reason to  $\Phi$  without this affecting their deliberations at all. But irrational people can think and do just about anything.

<sup>21</sup> What *else* could it consist in? Surely not in thinking that it is the kind of thing that people tend to take into account – this isn't a way of seeing something as a reason but a way of seeing something as seen as a reason.

<sup>22</sup> Zangwill also says, “motivational internalists intend to deny that there could be similar moral beliefs with different motivational profiles... The motivation debate is about whether belief determines or fixes

Zangwill claims that internalists who claim that motivation must be proportional to belief cannot account for accidie, depression and listlessness (2008a p. 55). In such cases moral motivation varies with no apparent variation in moral judgement. However, this poses no threat to the practical reason approach. We can see such cases as involving one or more of the following phenomena: an irrational inability to respond to recognised reasons, failure to recognise reasons as reasons, lack of engagement with the moral reasons on which the moral judgements are based or a lack of moral understanding. It is highly plausible that such cases display one or more of these features. They all seem to involve, for example, coming to no longer see things as important and reason-giving. Zangwill writes:

[Suppose two people] both believe that they should not accept a bribe but the first resists while the second does not. The second person takes the bribe, not because he irrationally succumbs to temptation, but because he cares less about morality than the first person, who refused the bribe (2008a p. 56).

Internalists should admit that people vary in the degree to which they care about morality. They might even admit that this might be due to differences in motivations that are external to judgements about moral reasons. Internalists claim that moral judgements can motivate without external help, not that they are the only source of motivation to act morally. However, internalists should also accept that moral judgements can vary between agents in their *intrinsic* motivational force without any implication that one of the agents is irrational. Irrationality can cause such variation, but another source of such variation can be found in varying degrees of moral understanding. If one sees moral considerations as strongly reason-giving (as, internalists claim, one should) then one will find them strongly intrinsically motivating. If one dismisses moral considerations as unimportant then they are unlikely to provide much motivation. That one can dismiss morality without irrationality does not mean that there is no *reason* why one shouldn't.

## 2.10 Amoralists and cynics

Zangwill argues that “the phenomenon of not caring very much about the demands of morality ... is best explained by externalism” (2008b p. 101). He gives real life

---

desire” (2008a p. 53). However, even if moral judgements are intrinsically motivating, there might be other sources of motivation that can combine with or compete with moral motivation.



examples of a coffee addict who cares less about everything (including morality) before his morning coffee, a mercenary who shows no moral motivation despite displaying moral knowledge, and a drunk driver who showed no remorse after his recklessness resulted in two deaths (2008b p. 102). These examples show that moral knowledge does not rule out moral indifference, or at least that it does not guarantee that one cares very much. Internalists should respond to such cases by maintaining that such agents either lack moral understanding (since they make moral judgements without understanding that the considerations on which such judgements are based constitute genuine reasons for action) or they are irrational (failing to act on reasons that they do recognise). Hence, internalists need to demonstrate that there are moral reasons.

Zangwill says that none of these cases necessarily involve irrationality (2008b p. 113). He writes:

It is beyond question that first thing in the morning, I have immediately present vivid desires for coffee. The question is: do I also have other less vivid desires that I am ignoring in my practical reasoning—so I am after all being irrational? (2008b p. 114)

Zangwill assumes an instrumentalist picture according to which practical reasoning must begin from one's desires. It will be unsurprising if morality is not required by instrumental rationality: it is unlikely that our reason for being moral is that it is a way of getting what we want. Internalists might deny that instrumental rationality is exhaustive of practical rationality and claim that therefore an agent might be instrumentally rational yet still fail to be practically rational. Zangwill thinks this move is ineffective since moral cynics might care about being instrumentally rational but not care about other norms of practical rationality any more than they care about morality (2008b p. 116). At least two responses are available to internalists. Firstly, internalists might claim that in fact instrumental norms requiring that one pursue the means to satisfying one's desires are not actually valid (see §§ 7.2-7.3). Secondly, the response that Zangwill puts into the mouth of the cynic might turn out to be irrelevant. We might be genuinely required to conform to norms of practical rationality whether we care about doing so or not. That said, when a cynic boasts "I don't give a damn and I don't see why I should" (Zangwill 2008b p. 120), internalists should not necessarily seek to characterise the cynic's mistake as involving irrationality. It may be, rather, that they are

failing to recognise a reason<sup>23</sup>. Again, the challenge to internalists is to show that there are reasons for everyone to be moral, whether these are supplied by norms of practical rationality or more directly by moral reasons<sup>24</sup>.

Internalists need to demonstrate that moral reasons are weighty, that morality is *important* and that it is generally far more important than (e.g.) fashion or etiquette, even if it is not always overriding. They cannot just complacently assume that moral judgements supply strong motivations to act morally. The facts tell us otherwise: many people “treat moral considerations as quite important as far as they go, but... they are more concerned with other things, such as their own interests, fashion, drugs, etiquette, or their children’s welfare” (Zangwill 2008b p. 106). Moreover, there are moral cynics who think morality is not worth caring about. Zangwill says that if such cynicism occurs and is “non-acratic and selfknowing, this gives us reason to believe that motivation does vary while moral belief remains constant” (2008b p. 108). Interestingly, Zangwill explains variation in moral motivation, at least in these cases, as explained by variation in normative beliefs: beliefs about what is worth caring about. If morality is worth caring about then it follows that only agents who are morally motivated get things right. This is the internalist claim: moral cynicism is a way of getting things wrong. The difficulty is in justifying this internalist claim: we need to demonstrate that there are moral reasons.

## 2.11 Internalism and Humeanism

The objections examined did not threaten the internalist thought that *if* there are moral reasons, the recognition of these reasons might be intrinsically motivating, and rational agents who display moral understanding might be motivated in accordance with their moral judgements. However, there remains a significant obstacle to establishing this form of internalism: it may be incompatible with the Humean theory of motivation. According to Smith, the Humean theory of motivation is the following claim:

---

<sup>23</sup> Externalists might claim that internalism is true only if we can *make cynics care*. This demand is unreasonable. We can offer reasons for caring but we cannot *force* the cynic to respond to them.

<sup>24</sup> Showing there is a reason to be moral might actually undermine internalism by supplying agents with an external motivation such that moral motivation is generated by judgements that something is morally required only when this is supplemented with the reason for caring about morality. This is a genuine threat, but whether it turns out to be worrying depends on the structure that a successful account of moral reasons has. An account could show us why moral considerations are genuine reasons without itself standing behind each of those considerations as its normative and motivating force.

An agent is motivated to act in a certain way just in case she has an appropriate desire and a means-end belief, where belief and desire are, in Hume's terms, distinct existences. (1994 p. 12, p. 126, p. 184)

This entails that no belief is intrinsically motivating. Moreover, according to Humeans, beliefs and desires are distinct and there are no necessary relations between them. For any belief-desire pair, it is possible to have the belief and lack the desire, or vice versa. Humeans define motivating reasons as follows:

Agent A at t has a motivating reason to  $\Phi$  only if there is some  $\psi$  such that, at t, A desires to  $\psi$  and believes that were she to  $\Phi$  she would  $\psi$ . (Smith 1994 p. 93)

According to Humeans, nothing is intrinsically motivating unless it is, or supplies, both a desire and an appropriately related means-end belief. However, internalists claim that recognising a moral reason can be intrinsically motivating. If recognising a moral reason is to be described in terms of beliefs and desires, it can only be described as *believing* that there is a moral reason. The problem is not merely that this is not a means-end belief. It is also that, since Humeans claim there are no necessary relations between beliefs and desires, no belief entails a desire. Humeans claim that (a) believing there is a reason does not entail the presence of desire and that (b) a desire is necessary for motivation. This entails that beliefs about reasons can never be intrinsically motivating.

On the internalist account, *her pain* could be my reason for administering a painkiller, and my recognition of her pain might explain my motivation and why I acted as I did. Humeans claim that this cannot be true because no belief can motivate except in combination with a desire. Humeans would have us explain the action by combination of a belief and desire, for example:

Belief: 'giving her a painkiller is a way to relieve her pain'

Desire: 'to relieve her pain'.

Internalists claim that the belief in this pair could itself be intrinsically motivating. Positing the desire as well makes the motivation external to the judgement. Internalists may accept that motivation may sometimes be external in this way, but they deny that it must be. Since Humeans claim that motivation is necessarily external to judgement, the Humean theory is incompatible with internalism.

Smith thinks internalism can be reconciled with the Humean theory. Smith says that when an agent comes to believe a moral judgement of the form ‘it is right that I  $\Phi$ ’ this expresses a belief about what they have normative reason to do (1994 p. 185) which, in turn, causes the agent, if rational, to come to desire to  $\Phi$ . Therefore, in rational agents the judgement that it is right to  $\Phi$  is always accompanied by a desire to  $\Phi$ . This seems straightforwardly inconsistent with the Humean claim that there are no necessary relations between beliefs and desires. Let us ignore this point, however<sup>25</sup>. If the agent has some idea of how to go about  $\Phi$ -ing then the agent has some belief that  $\psi$ -ing is a way to  $\Phi$ . According to Humeans, this belief together with the desire to  $\Phi$  constitutes a motivating reason to  $\psi$ . This explains the agent’s  $\psi$ -ing<sup>26</sup>. Hence, we have a Humean explanation of an action that was motivated by a moral judgement without the input of any external desire or motivational state.

However, we wanted a Humean-internalist explanation of how the agent’s judgement that it was right to  $\Phi$  explains their  $\Phi$ -ing. We have only shown how it explains their  $\psi$ -ing. It might be suggested that once we have explained their  $\psi$ -ing we have explained their  $\Phi$ -ing, since by  $\psi$ -ing they  $\Phi$ -ed. This, however, will not do. Suppose I judge that it is right for me to give £100 to Oxfam. Suppose that I make this judgement because I judge that giving £100 to Oxfam will save a child’s life and that this is a normative reason for me to give. Suppose that I am rational, and that my appreciation of this normative reason leads me to desire that I give £100 to Oxfam. Suppose also that I believe that in order for me to give £100 to Oxfam I must sell my watch. If I sell my watch we have a Humean explanation of why I did so: I desired to give £100 to Oxfam and believed that by selling my watch I could do so. However, if after I sell my watch I give £100 to Oxfam, we have no Humean-internalist explanation of this. We have an explanation all right: I gave £100 to Oxfam because I judged that I had a good reason to do so and that it was right for me to do so. According to Humeans, this is not really an explanation of my action; or at least, not until something else is added to it. What else could we add? That I wanted to do what I had judged there was a reason to do? That I wanted to do what was right? Neither of these options is appealing and neither is consistent with internalism.

---

<sup>25</sup> Perhaps this is a *normative* relation since it is necessary only in rational agents. It is unclear whether this is consistent with the Humean theory or not.

<sup>26</sup> If the judgement that it is right to  $\Phi$  is correct then it also justifies  $\psi$ -ing.

Perhaps we are overlooking the most obvious possibility here: that being rational is a matter of having certain desires that ensure that one is motivated to do what one believes one has reason to do. This would explain why rational agents are necessarily motivated to do what they believe they have reason to do, but irrational agents are not. However, this is still a form of externalism: the motivational force of reasons and moral judgements is supplied *externally* by an additional desire. Once again, the Humean theory drives us back to externalism.

## 2.12 Conclusions

An appealing form of moral internalism has been outlined according to which rational agents are motivated by moral reasons that they recognise and rational agents are motivated in accordance with their moral judgements insofar as they display moral understanding. It has been claimed that this form of internalism would be vindicated if we can establish that there are moral reasons. However, it has been seen that this form of internalism is incompatible with the Humean theory of motivation. In the chapters that follow (three to six) I examine the Humean theory and conclude that we should reject it.

## Chapter 3. Commonsense and the Humean theory

### 3.0 Introduction

Our attempts to understand morality as essentially practical were frustrated by the Humean theory of motivation (hereafter, simply ‘Humean theory’). We have been working with a definition of the Humean theory provided by Smith (1987, 1988a, 1998b, 1994, 2003) which, being the best defended, remains the focus in later chapters (five and six). However, other related Humean claims are also considered. In this chapter I claim that the Humean theory tends to be assumed rather than argued for (§ 3.1), and set out various Humean claims (§ 3.2) and conceptions of desire (§ 3.3). In the bulk of the chapter (§§ 3.4-3.11) I argue that the Humean theory is not commonsense. I aim to establish that the Humean theory gains no direct support from commonsense and requires philosophical support that it is rarely given. This lays the foundations for later chapters (four to six), where I examine arguments for the Humean theory.

### 3.1 The Humean theory: admittedly a dogma

Broadly speaking, the Humean theory is the idea that since actions are intentional or performed for reasons, they are to be explained in terms of the reason why the person acted, and that giving such a reason explanation always involves mentioning a belief-desire pair. Simple examples are often given to illustrate this idea: “you cross the street, because you believe that there is a florist’s shop on the other side of the street, and you want to visit a florist’s shop” (Wedgwood 2002 p. 345).

Some might think it unnecessary to argue that the Humean theory is not commonsense, since it is obviously a philosophical theory. However, others may find it equally obvious not only that the Humean theory *is* commonsense, but also that it is true. Furthermore, whilst the Humean theory is almost always treated as a philosophical theory in the metaethical literature, it is often treated as the default view, a theory that is obviously true and requires only the support of intuitions about particular examples such as the florist’s shop example above. The fact that the theory is so often asserted and so rarely defended seems to indicate that it is presumed to be commonsense. It is introduced in the context of commonsensical examples and then assumed that, since no

one can dispute with its verdicts about these examples, it is an enlightening explanatory generalisation. The following passage from Zangwill (2008b) is representative of these tendencies:

[I]t is common sense that we do often appeal to distinct beliefs and desires in order to explain and predict actions. The particular roles that beliefs and desires play in much of our common-sense explanatory practice, and in all of it according to Humeans, is that they are thought of as distinct states that mesh together in a particular way to yield action: beliefs function to supply the means to the end specified in the content of desires, or they inform us of the existence of what we desire. So distinct beliefs and desires are each necessary for motivation; neither can motivate without the other. Only together can they motivate. To use a couple of old examples: I believe that if I step on the ice I will fall through and become wet and cold. I do not want to become wet and cold. These two states can motivate me not to step on the ice. But the belief that the ice is thin and that I will fall through if I step on it will only motivate me not to step onto the ice if I don't want to get wet and cold. If I liked being wet and cold, I would no doubt leap onto the ice with abandon. (2008b p. 98)

Importantly, since the inference from 'we do often appeal to distinct beliefs and desires' to 'so beliefs and desires are each necessary for motivation' would represent an obvious non-sequitur, Zangwill cannot plausibly be interpreted as offering an *argument* for the Humean theory here. The intervening claim that beliefs and desires are 'distinct states that mesh together in a particular way to yield action' is not a step in an argument but a statement of the claim to be established. Moreover, the examples that follow are hardly sufficient grounds for generalising to all actions, and it is not uncontroversial as to what examples of this kind actually show. If Zangwill is not offering us an argument for the Humean theory, his claim must be either that the truth of the Humean theory is commonsense<sup>1</sup> or that the Humean theory is a way of formalising some of our commonsense ideas<sup>2</sup>. If either of these claims is right it will provide substantial support for the Humean theory. It is therefore necessary for me to show that neither of these claims is right.

Another possibility is that authors such as Zangwill adopt the Humean theory of motivation by a process of elimination. Mele, for example, assumes, without providing

---

<sup>1</sup> I think that in fact Zangwill takes himself to be both explaining and supporting what he takes to be the standard model of motivation by appeal to commonsense and everyday intuitions. The Humean theory *is* the standard model of motivation (cf. Audi 1990 p. 420, Brandt 1990 pp. 404-5, Harrison 1979 p. vii, Pettit and Smith 1990 p. 565) but this doesn't mean that it is correct or that it doesn't need support.

<sup>2</sup> A third possibility is that he might be thinking of the Humean theory as a well established philosophical theory, and have some particular arguments in mind that establish it. This possibility is considered in chapters four to six.

any positive support for the claim, “that all motivation-constituting items are representational states of mind – or ‘attitudes,’ for short” (1995 p. 388) and that all intentional actions have among their causes “mental items (events or states), including motivation-constituting attitudes” (*ibid.*). He only considers three alternatives to this view. The first is that intentional actions are not caused, and are to be explained non-causally in terms of mental items (see, e.g., Melden 1961). The second is that intentional actions are caused but admit of non-causal explanations in terms of mental items (see, e.g., Peters 1958, Wilson 1989). The third is behaviourism: that intentional actions are caused but mental states do not play any role in any valid explanation of them (Skinner 1953). The assumption seems to be that the only candidates for explanation of action are either reductive scientific explanations, explanations in terms of beliefs and desires, or a combination thereof. Pettit and Smith (1990 p. 565) also assume that those philosophers who do not accept that actions must have belief-desire explanations are those who want to replace it with rigorous cognitive science, such as Churchland (1979, 1981) and Stich (1983). I argue that there is neither reason to accept this assumption nor reason to think that the Humean theory is supported by commonsense. However, before we move on to this task let us get clear about what the Humean theory is in order that we may see clearly whether commonsense supports it or not.

### 3.2 Some Humean claims

The Humean theory can be characterised in a number of different ways. Humeans are said to claim that ‘reason alone cannot motivate’, that ‘reason is the slave of the passions’, that desires must feature in explanations of action, or that motivating reasons are constituted by belief-desire pairs. Those who call themselves Humeans do not necessarily defend all the claims that might be labelled ‘Humean’. Smith (1988a), for example, defends both moral rationalism and the Humean theory, taking the issue that divides Humeans and non-Humeans to be distinct from that dividing rationalists and anti-rationalists. He claims that motivating reasons are constituted by belief-desire pairs, but denies that reason is the slave of the passions. Although it is unclear whether Smith’s attempt to combine rationalism with the Humean theory is successful, it highlights that the two positions are not straightforwardly contradictory. With this in



mind we should distinguish between a number of related claims that Humeans might make:

1. *The teleological conception of action*: Since intentional action is purposive, it is goal directed in the sense that it aims at some end state or outcome.
2. *Psychologism about motivating reasons*: Giving the agent's reason for action always involves mentioning one or more of the agent's mental states.
3. *The belief-desire thesis*: A belief-desire pair is required to motivate or explain an action; the belief in the pair is a belief that the action in question will contribute to the satisfaction of the desire in the pair. Such a belief-desire pair constitutes the agent's motivating reason.
4. *Belief-desire as psychologically real*: In attributing a belief-desire pair we attribute to the agent two independent psychologically real mental states. The belief-desire explanation of the agent's action succeeds only if the agent genuinely possessed both of these mental states at the time of action or decision.
5. *Belief-desire distinctness*: Desires and beliefs are distinct existences. There are no necessary relations between them, and there are no states that are belief-like in one respect but desire-like in another.
6. *Direction of fit*: Beliefs and desires are to be distinguished by their directions of fit with the world: beliefs fit to the world, the world fits to desires.
7. *The means-end rationality assumption*: Belief-desire pairs only explain actions on the assumption that the agent is practically rational or means-end rational. Being means-end rational is a matter of doing what one believes will contribute to the satisfaction of the desires that one has.
8. *Desire as non-rational*: Desires are not rationally assessable except to the extent that they are based on false beliefs.

9. *Insufficiency of normative explanations*: Citing a norm that directs an agent to action is not sufficient to explain that action. To explain the action with respect to the norm, the agent must have a belief that the norm requires the action and a desire to do what the norm requires.
10. *Reason as inert*: Since rational principles alone cannot direct action, reason is motivationally inert.

These claims all relate to the idea that when an agent acts intentionally, they act for a reason, the action is *theirs* and has its *source* in their agency: the source of action is internal rather than external to the agent. Humeans account for this by locating the source of action in the agent's mental states, thus providing a way of distinguishing things that a particular agent *does* from all other kinds of event that may involve that agent.

We can distinguish between metaphysical and explanatory interpretations of Humean claims such as the belief-desire thesis. On the metaphysical interpretation, the belief-desire thesis is the following *ontological* claim about the *essence* of action: action *essentially* involves the *presence* of paired beliefs and desires. On the explanatory interpretation, the belief-desire thesis only tells us what form action *explanations*<sup>3</sup> must take. This does not tell us what the *essence* of actions is or whether the mental states mentioned must actually be *present*<sup>4</sup> in order for the explanation to be valid and the event to count as an action. Hence, the explanatory belief-desire thesis does not entail the ontological belief-desire thesis. Conversely, the ontological belief-desire thesis does not entail the explanatory belief-desire thesis since it might be that belief-desire pairs are always present in cases of action, but that mentioning them does not help to explain actions. The distinction between the metaphysical and explanatory claims is not often mentioned, although Ruben (1997) alludes to it. In discussing the claim that belief-desire pairs cause actions, he notes that this is “a metaphysical or ontological view about what a token action is, and is silent ... on the question of the explanation of action” (1997 p. 230).

---

<sup>3</sup> If it is thought that an ‘action explanation’ is necessarily a description of that action then something that motivates an action might only be *part* of an action explanation. Humeans, however, think that what motivates an action always contains a description of the action in the content of the agent's belief. For the Humean, an action is always motivated under a particular description, and hence mentioning the belief-desire pair that motivated the action describes the action and provides an ‘action explanation’.

Different anti-Humean authors vary in their identification of the central Humean claim to be opposed. For now I leave aside this issue, although ultimately I will argue that we have no reason to accept any of the Humean claims listed above (1-10), whether taken individually or together<sup>5</sup>. In this chapter, since I discuss whether the Humean claims are commonsense, I focus mostly on the belief-desire thesis (3) and that which it presupposes, psychologism about motivating reasons (2). I also have something to say about the insufficiency of normative explanations (9), since it relates closely to the belief-desire thesis. It is the belief-desire thesis that is widely claimed to be commonsensical. The other claims are too technical to be plausibly construed as commonsense, and they are dealt with in other chapters. Therefore, in this chapter my main claim is that there is no reason to accept that commonsense explanation of action takes the belief-desire form. I claim that, on the contrary, we should allow that explanations of action are multifarious, and do not always involve mentioning the agent's reason for action. I claim that when we do give an agent's reason for action we do not ordinarily give a belief-desire pair, but aim to show the favourable light in which the agent saw the action. Doing this might take any number of forms not captured by the belief-desire thesis.

### 3.3 Seven conceptions of desire

Many of the claims (1-10) above involve the term 'desire' in a crucial way: whether each claim is true depends on what exactly is meant by 'desire'. It cannot be assumed that the meaning of 'desire' is clearly understood, since there are a number of incompatible understandings of this term. These will be discussed later (chapters four and five); for now I just wish to point out that these different conceptions are available, since much confusion can result from conflating them. I distinguish between seven conceptions of desire:

- (a) **The ordinary conception of desire:** The ordinary conception of desire is simply whatever reports our everyday usage of and linguistic intuitions about the term

---

<sup>4</sup> Or, we might say, the agent must actually *have* or *be in* those states.

<sup>5</sup> I say in chapter four that there is one interpretation, suggested by Davidson, under which the belief-desire thesis (3) is true. But under that interpretation it is strange and misleading to use the word 'desire', and it neither presupposes nor entails any of the other claims (1-2, 4-10).

‘desire’. This may or may not be accurately reported by one of the other conceptions of desire listed below.

- (b) **The pro attitude conception of desire:** According to the pro attitude conception of desire, anything that could motivate an agent to action should be referred to as a desire. Pro attitudes are discussed at length in chapter four.
- (c) **The ‘direction of fit’ conception of desire:** The ‘direction of fit’ conception of desire is a technical notion of desire that claims that desires are states with which the world must fit. In chapter five I discuss this notion and conclude that we have no reason to accept this conception of desire or the Humean picture it supposedly supports.
- (d) **The functional or dispositional conception of desire:** This conception of desire is a development of the ‘direction of fit’ conception of desire. It is proposed that we can understand desires as dispositions analysed in terms of conditionals. I consider and reject this idea in chapter five.
- (e) **The normative conception of desire:** According to the normative conception of desire, desires are states that give us (good or normative) reasons for action. This conception is discussed and rejected in chapters six and seven.
- (f) **The phenomenological conception of desire:** According to the phenomenological conception of desire, desires are states that essentially have phenomenological content. This conception is discussed in chapter four.
- (g) **Proper desires:** Schueler (1995) claims that the ordinary sense of desire, as opposed to the philosopher’s (pro attitude) conception, is captured by ‘proper desires’. For Schueler, the defining feature of proper desires is that it is possible for people to do what they have no proper desire to do.

I discuss ‘proper desires’ in chapter four. Importantly, simply asserting without argument, as Schueler seems to, that our ordinary conception of desire is that of ‘proper desires’ comes close to begging the question against at least one kind of Humean claim.

The point needs to be substantiated with argument and illustrated with examples. However, my aim in the rest of this chapter is to achieve something of this type. That is, I aim to show that commonsense conceptions, explanations and predictions of actions do not necessarily or significantly involve the ascription of belief-desire pairs.

### 3.4 Commonsense and the Humean theory

It is often supposed that nothing is an action unless it is something the agent did intentionally. It is then added that anything that an agent did intentionally, they did for a reason, and this ‘motivating reason’ explains the action. Therefore, it is thought, any action is done for a motivating reason. According to the Humean theory, a motivating reason is an appropriate belief-desire pair. A belief-desire pair is appropriate when the belief in the pair is a belief about how the action in question would satisfy or contribute towards satisfying the desire in the pair. If the Humean theory is true then for any action there is a belief-desire pair that, by constituting the agent’s motivating reason, explains the action. Hence, actions are *rationalised* by belief-desire pairs.

Although plausible, the claim that all actions are motivated by reasons might be rejected. If it were, Humeans would have to restrict their claim to the subset of actions motivated by reasons. This possibility is not explicitly discussed here, but I argue that our everyday practice of explaining actions, whether in terms of reasons or not, is not captured by the Humean belief-desire formula. It is worth noting that the Humean theory can be characterised as the claim that attribution of a belief-desire pair is both necessary and sufficient<sup>6</sup> for explaining an action. It is possible therefore to split the question of whether the Humean theory is commonsense into two further questions:

---

<sup>6</sup> We must distinguish between two kinds of sufficiency here. In the first sense, the sufficiency of belief-desire explanations means that once we have mentioned an appropriate belief-desire pair, there is never anything else that we must *add* in order to complete the explanation. In the second sense, if a belief-desire pair is mentioned this *guarantees* that the action is adequately explained. The second sense of sufficiency is too strong for the Humean. A belief-desire attribution cannot guarantee an adequate explanation, because adding something else might ruin the explanation. For example, mentioning that I desire ice cream and believe there is ice cream in the freezer might explain my going to the freezer. But adding that I believe eating peanuts will kill me, that I believe that the ice cream contains peanuts and that I desire to stay alive ruins the original explanation. We can call such additions ‘explanatory disablers’, since they resemble the kind of disablers discussed by Dancy (2004 p. 41) in relation to normative reasons. Since the Humean wants to maintain that mentioning the belief-desire pair is what explains the action, they want to deny that any other explanatory factor is needed. They want to deny that explanations must mention the absence of all the beliefs and desires that the agent lacked that, if present, would have ruined the explanation. That is, they do not want to have to give ‘full explanations’ of the type endorsed by Crisp (2000 p. 44, cf. Dancy 2004 pp. 46-7). So the Humean only claims belief-desire explanations are sufficient in the first sense. (Even then, they sometimes add that such belief-desire pairs only explain on

- (i) Are belief-desire attributions necessary for giving commonsense explanations of actions, or are there commonsense explanations not given in terms of beliefs and desires?
- (ii) Is a belief-desire attribution sufficient for giving a commonsense explanation of an action, or do we sometimes need to mention additional facts?

The answers to these questions do not admit of degrees: either something is necessary or it is not; either it is sufficient or it is not. However, since we are talking about commonsense, perhaps we should not expect strict necessity or sufficiency. The necessity or sufficiency of these explanations might hold for the most part, allowing exceptions or other forms of explanation in some cases. However, if the Humean theory is based in commonsense, these exceptions must not be too numerous. They should not indicate that there is any other important and interesting way of explaining human action.

### 3.5 The Humean theory: a ‘folk psychology’?

The connections between folk psychology and the Humean theory are not often discussed. References to folk psychology are typically found in literature on cognitive science and philosophy of mind whereas the Humean theory is usually discussed in relation to moral psychology and practical reason. However, the links are remarkably close. Proponents of folk psychology typically view the kind of belief-desire rationalisations of actions that Humeans favour as being paradigmatic of our ordinary interpersonal understanding.

Ratcliffe (2007b p. 224) and Ratcliffe and Hutto (2007 p. 2 cf. Davies and Stone 1995 p. 2; Kusch 2007 p. 175) define folk psychology as claiming:

- (1) Everyday interpersonal understanding is facilitated by a ‘commonsense’ or ‘folk’ psychology
- (2) We employ this folk psychology to make sense of actions by interpreting them in terms of reasons composed of propositional or intentional attitudes: beliefs and desires

---

the assumption that the agent is practically rational or means-end rational. But for now we can ignore this

- (3) Our ability to predict and explain actions relies on this folk-psychological activity

According to Ratcliffe and Hutto (2007 p. 1), proponents of folk psychology characterise beliefs as states that “carry information about the world and thus guide action” and desires as “motivational states that specify goals for actions”.

Proponents of folk psychology, such as Stich and Ravenscroft (1994), are in agreement with this characterisation. They say, “[w]e use terms like ‘believe’, ‘think’, ‘want’ and ‘desire’ to describe ourselves and each other” (1994 p. 457), and this “is a manifestation of a widely shared capacity to describe ... people using intentional idioms” (*ibid.*). They claim that we use these descriptions “to construct explanations of people’s behavior” (*ibid.*) utilising ‘because’ clauses, and that we can produce predictions in the same manner. Hence they affirm all three of the claims above. They also claim that “[t]here is a wide range of generalizations about the interactions among stimuli, mental states and behavior that people in our culture occasionally utter, and are generally quite willing to endorse when asked” (p. 458). Stich and Ravenscroft (1994 p. 458) present the following examples:

When a normal person is looking at a traffic light which changes from red to green she usually comes to believe that it has changed from red to green.

If a person believes that all scorpions are poisonous, and if she comes to believe that Henry’s pet is a scorpion, then she will typically come to believe that Henry’s pet is poisonous.

If a person sitting at a bar wants to order a beer, and if she has no stronger desire to do something that is incompatible with ordering a beer, then typically she will order a beer.

They assume that since no one would ordinarily deny that these statements are true, these statements play an important role in our everyday interpersonal understanding.

Some proponents of folk psychology endorse ‘theory theory’ (Morton 1970) whilst others favour ‘simulation theory’ (see Davies and Stone 1995). According to ‘theory theory’, interpersonal understanding is facilitated by a theory, a body of theoretical knowledge, which the ‘folk’ apply to particular cases. Of ‘theory theory’, Kusch writes:

[A]llegedly a network of laws enables its users to explain and predict observable behaviours by postulating unobservable ‘theoretical entities’. In the case of folk psychology, these laws are platitudes like ‘someone who is thirsty desires to drink’, or ‘someone who believes herself successful is likely to feel pride in her achievement’. The theoretical entities are, first and foremost, beliefs and desires. (2007 pp. 175-6)

The simulation theory claims that these achievements in prediction and explanation, relating probable beliefs, desires and actions, are not produced by applying theoretical knowledge, but by undergoing imaginative processes of simulation, similar to one’s own deliberations about what to do (see e.g. Stich and Ravenscroft 1994 p. 464).

My discussion will only seek to discredit ‘theory theory’ and ‘simulation theory’ to the extent that they might resemble or support the Humean theory. Since ‘theory theory’ is the more similar to the Humean theory, it will receive greater attention. Before proceeding, let us summarise the claims of the ‘theory theory’ of folk psychology. They are listed in increasing strength, and each claim should be interpreted so that it presupposes and includes all the previous claims in the list:

1. There is a commonsense method that all people ordinarily use for understanding, predicting and explaining each other’s behaviour.
2. This method involves (perhaps tacitly or perhaps explicitly) employing a theory.
3. The commonsense theory that facilitates interpersonal understanding involves attributing mental states to people in order to predict and explain their behaviour.<sup>7</sup>
4. The mental states attributed in prediction and explanation are divisible into two fundamental types, beliefs and desires.
5. A prediction or explanation of a piece of behaviour is not adequate unless it mentions beliefs and desires. The presence of either a relevant belief or desire can be implicit, but it must be possible to make it explicit, otherwise the prediction or explanation is not adequate.
6. Each piece of behaviour can only be explained by mentioning both a desire, and a belief that the behaviour is a way of satisfying, or contributing to the satisfaction of, that desire.



These claims have been given in terms of ‘behaviour’. However, each of them could also be made about actions, voluntary actions, intentional actions, or actions that are done for a reason. Even if the claims (1-6) were not generally true of behaviour, they could be generally true of some sub-domain of behaviour, such as actions done for a reason. Showing that the Humean theory gains no support from commonsense requires showing that the claims are not universally true in any of these domains.

The argument in this chapter will not entail anything about the truth or falsity of the first of these claims (1), but it is claimed that all of the other claims are false when interpreted as claims about explaining and predicting<sup>8</sup> behaviour. Moreover, it is claimed that they are also false when construed more narrowly as claims about action for a reason.

### 3.6 Commonsense explanations

Those who discuss folk psychology “generally take the existence and ubiquity of folk psychology (construed as the attribution of internal propositional attitudes) for granted” (Ratcliffe and Hutto 2007 p.2). Argument or empirical evidence is not presented in favour of folk psychology, because, since it is commonsense, it is assumed that it will be recognised as such. However, examples are often offered. Consider again one of the examples offered by Stich and Ravenscroft:

If a person sitting at a bar wants to order a beer, and if she has no stronger desire to do something that is incompatible with ordering a beer, then typically she will order a beer. (1994 p. 458)

After giving three such examples, they remark:

We trust you agreed with all of them. In so doing you were manifesting the widely shared ability to recognize folk psychological generalizations (*ibid.*).

---

<sup>7</sup> This is termed ‘theoretical’ since the posited mental states are unobserved states attributed on the basis of observed events. The idea is that they are like theoretical entities posited in science (e.g. electrons) in this respect.

<sup>8</sup> Prediction is not discussed explicitly because of limitations in space. It is less important for the purposes of assessing the Humean theory since the Humean theory is a thesis about explanation. It is assumed here that much the same considerations apply to prediction as to explanation. However, if they do not then this should not affect my conclusions about the Humean theory.

Unfortunately for Stich and Ravenscroft, this is a non-sequitur. The fact that we were unwilling to deny the example statement does not entail that it expresses something that we would ordinarily assert or think commonsensical. In fact, I think the claim given in Stich and Ravenscroft's example would strike most of the 'folk' as being distinctly odd, and not the kind of thing they would ordinarily assert<sup>9</sup>, even if it were true. Were we to accept that this assertion, and others of its kind, are commonsensical, we would not thereby be accepting the further claim that this kind of assertion is of a form uniquely useful in explaining and predicting human behaviour. In fact, the proposition in question actually seems particularly badly suited both to predicting and to explaining, regardless of whether the proposition is given as an explanation itself or used to facilitate other explanations. Typically, reporting that someone wanted to order a beer more than she wanted to do anything else that was incompatible with it tells us nothing at all, simply because it does not rule anything out. If people always do what they most want to do then the sense of 'want' in question must be of the broadest kind that includes anything that might motivate someone. In this sense, saying that someone did something 'because they most wanted to' tells us nothing about *why* the person did it; it only tells us *that* they did it<sup>10</sup>. We can conclude then that the mere presentation of examples of this kind is not only insufficient to establish folk psychology but provides it with no support at all unless it is back up with arguments or evidence. Are such arguments and evidence forthcoming?

So far it seems that no support for folk psychology has been provided by scientific studies. Scientific studies on action explanation and prediction, such as those investigating whether young children can take false beliefs into account in predicting behaviour (Wimmer and Perner 1983) have *presupposed* that belief-desire explanations of action are commonsensical (Ratcliffe 2007b p. 229) rather than providing support for this claim. Ratcliffe claims that "our most deeply entrenched concepts will most likely be presupposed as a lens through which scientific results are interpreted, rather than discovered through them" (2007b 230). Whatever our everyday way of understanding

---

<sup>9</sup> It might be suggested that whilst people would not *say* things like this it nevertheless this captures what they are thinking. However, it doesn't seem that we ever explicitly entertain such thoughts except, perhaps, when doing philosophy. It is true that we may sometimes think 'well if he wants to order a beer he'll probably order a beer', but only in particular contexts (e.g. 'he doesn't care about anyone else, he just does what he wants, so if he wants to order a beer...'), and not in the form universal form suggested by Stich and Ravenscroft.

<sup>10</sup> Or perhaps, that they did it without being forced or compelled to. In cases where one is forced or compelled Humeans will say that either one must have wanted to do it *in some sense* (e.g. to avoid being killed) or one didn't really do it at all (e.g. when someone is dragged into a prison cell).

others is, it will be as much taken for granted by scientific researchers as by the rest of us. Furthermore, researchers investigating interpersonal understanding are likely to already have substantial theoretical assumptions about its nature. If these theoretical assumptions are presupposed in designing the experiment and used to interpret the results, the results are likely to indicate that they represent our commonsense modes of interpersonal understanding. Hence, without the necessary prior philosophical reflection, scientific experiments may present a confused picture of commonsense psychology.

Ravenscroft, in assessing the relative merits of simulation theory and theory theory, takes the Humean theory as the default position. Presumably he takes it to be either commonsense or well established by philosophical argument. However, he makes no reference to any argument that he takes the view to be established by. The most natural interpretation of what Ravenscroft says is that he thinks the Humean theory was established by Hume himself:

So far we have been merely assuming that practical reasoning is Humean; that is, we have simply assumed that practical reasoning involves both beliefs and desires (Hume, 1739)<sup>11</sup>. For example:

1. I desire a beer.
  2. I believe that if I go to the fridge I can get a beer.
- Therefore
3. I will go to the fridge.

However, anti-Humean theories of practical reasoning exist. Thus, the example just given might be reconstructed as follows:

- 1'. I believe that it would be prudent/moral/rational to have a beer.
  2. I believe that if I go to the fridge I can get a beer.
- Therefore
3. I will go to the fridge. (Ravenscroft 2003 p. 170)

Ravenscroft considers two arguments for the anti-Humean conception, and concludes that neither gives us reason to reject the Humean conception. However, the Humean and anti-Humean conceptions that Ravenscroft presents are not exhaustive – they might both be false. Therefore, arguments against the anti-Humean theory are insufficient to establish the Humean theory. Moreover, Ravenscroft never gives us any positive reason

---

<sup>11</sup> This is the reference given by Ravenscroft to Hume's *A Treatise of Human Nature*. I have left it in the quotation to illustrate that Ravenscroft appears to accept the theory on Hume's authority, although without referring to any specific passage in the *Treatise*.

for accepting the Humean conception: he admits that we have been ‘merely assuming’ it and yet still fails to provide any positive arguments. Inserting a reference to Hume does not suffice<sup>12</sup>. Importantly, it is never even stated what it would mean for practical reasoning to be Humean. Would that mean that instances of reasoning that do not conform to this pattern are (a) irrational (b), cannot issue in action or (c) fall outside the domain of practical reasoning? In the context of a discussion of theory theory and folk psychology, we also want to know what the connection is between practical reasoning and commonsense explanations of action.<sup>13</sup> Asserting that practical reasoning is Humean is not the same thing as asserting that explanations of actions must be Humean. We can conclude, therefore, that Ravenscroft offers us no reason whatsoever to accept folk psychology or the Humean theory.

### 3.7 Ordinary reasons explanations

Hutto claims that “if the action was performed for [a] reason only explanations that bring the folk psychological framework will do” (2007 p. 130 fn. 7). He claims that “folk psychologists trade in *reasons* – minimally, belief/desire pairs – not just isolated thoughts, or desires” (Hutto 2007 p. 119). But Hutto just seems to assume that reasons must be analysed as belief/desire pairs, rather than providing any argument for the claim. Hutto does illustrate what he means, however, in the following way:

He left the party because he believed the host had insulted him. She will head for the cabin in the woods because she wants peace and quiet. These are typical examples of reasons explanations... To leave a party because of a suspected insult suggests that one desires not to be insulted, or at least that the desire to avoid insult is stronger than that for some other good on offer. Similarly, to seek tranquillity in an isolated cabin implies that one believes that it can be found there, or at least more so than elsewhere. (Hutto 2007 p. 122)

---

<sup>12</sup> We shall see in chapter four that the arguments Hume himself gives do not establish the Humean theory.

<sup>13</sup> It seems that Ravenscroft conflates these issues, for he discusses the question of practical reasoning in relation to *explanatory* theories of action. We have never been offered any reason to think that practical reasoning typically involves beginning from one’s mental states. In fact, it seems utterly bizarre to move from observations about one’s own psychology to a decision to act in a particular way. The first two premises in Ravenscroft’s inferences are observations, observations that a particular agent – myself – has particular mental states, and the conclusion is something utterly different in type – an action, an intention, or at least a judgement about what should be done. (Presumably the conclusion is not meant as a prediction, otherwise the reasoning would be theoretical, not practical). This is not (yet) a criticism of the Humean theory, but of Ravenscroft’s conflation of the deliberative and explanatory standpoints. The Humean theory is concerned with explanation, not deliberation.

However it is far from clear as to whether these examples illustrate Hutto's point. Leaving a party because one has been insulted is not best explained by a desire to avoid further insults. Leaving can be understood as a *reaction* to what has occurred, and this is best explained by one's emotional state, not by a desire. The Humean theory will struggle to deal with any analogous case, any case of reaction, in which the action is explained by reference to the *past* rather than the future. Desires that are posited in Humean explanations of action must be future looking, since the content of a pre-existing desire is supposed to explain an action that occurs at a *later* time<sup>14</sup>. Humeans cannot straightforwardly make sense of e.g. leaving a party as a *reaction* to being insulted. They need to posit an additional desire – to avoid being insulted again, or to show their displeasure – if they are to explain the action. This is a weakness of the Humean theory, because it rules out something that seems quite possible – that the insulted person had no desire that explained their action. Rather, their action was explained by their being upset or offended. Since commonsense would seem not to dictate that they must have had some additional desire, the Humean theory seems to be in conflict with commonsense.

Actions that are explained by emotions also seem problematic for the Humean theory. We might think, as Ratcliffe does, that whilst emotions are intentional states, they are not propositional attitudes<sup>15</sup>; “[a]n emotion is, among other things, a way in which an entity or situation is experienced” (Ratcliffe 2007a p. 198). Since a way of experiencing something is not reducible to beliefs and desires, emotions are not reducible to beliefs and desires<sup>16</sup>. Yet they play important roles in interpersonal understanding and explanation of action. This is obscured by insisting that our everyday explanations must be in terms of propositional attitudes such as beliefs and desires.

Why, then, does Hutto assume that actions performed for reasons are to be explained by belief-desire pairs? Interestingly, he writes:

---

<sup>14</sup> It is possible, of course, to desire that things had been different in the past. But Humeans do not typically appeal to such desires to explain action. Humeans claim that a desire explains an action when the content of the desire matches with some state of the world that can be brought about by the action. Since a desire that things had been different is a desire for something to occur (or not occur, or occur differently) in the past, and no actions can make something occur in the past, no desire that things had been differently can explain an action.

<sup>15</sup> Whilst this point is controversial, the controversy is problematic for the Humean theory and not for a permissive anti-Humean theory: Humeans need to show that actions explained by emotions must be explained in terms of beliefs and desires. Humeans therefore need a theory that explains emotions in terms of beliefs and desires. A permissive anti-Humean theory is compatible with *any* theory of the emotions.

In speaking of ‘reasons’ I mean precisely what philosophers have long understood to be the heart of discrete episodes of means-end practical reasoning – processes that result in intentions to act. (2007 p. 115)

It is unclear whether Hutto thinks that the Humean theory is commonsense after all: does he think that it is a theory *about* commonsense that draws support from a philosophical theory of practical reasoning? Or is it (also?) a philosophical theory drawing support from commonsense? Either way, we have not uncovered any positive reason for thinking that commonsense action explanations must always take the form of belief-desire attributions. We have also found at least two reasons for thinking that they do not: (1) explanations of actions that are *reactions* do not involve desires; (2) explanations of actions in terms of emotions cannot be reduced to belief-desire explanations. However, it might be thought that reactions only fail to involve desires because they are not motivated by reasons. Similarly, it might be that emotions are not reasons, so any action motivated by emotion is not done for a reason. It might be claimed that, even if reactions and emotion-based actions do not fall under the Humean theory, all actions done for reasons do. Therefore, I now consider explanations in terms of reasons. In the next section (3.8) I consider the possibility that the importance of reasons explanations may not be culturally universal. Having overcome this worry, I argue (3.10) that reasons explanations do not necessarily take the belief-desire form.

### 3.8 Reasons explanations and cultural variation

Hutto claims that “it is not a given that all cultures engage in the practice of understanding actions in terms of reasons” (2007 p. 116). He bases this claim on cross-cultural studies on belief and representation (Vinden 1996); judgements about beliefs and desires (Vinden 1999); the relative prevalence of dispositional, trait-based, contextual, personality-based, and spiritual explanations (Lillard 1997); and understandings of the nature of, and relation between, mind and behaviour (Lillard 1998). It is important to be clear about the different ideas in play here, however. Until it is established that reasons-talk must be universally understood in terms of belief-desire-talk, it cannot be assumed that findings about the latter are findings about the former. Admittedly, the studies cited by Hutto may indicate cultural differences in how actions

---

<sup>16</sup> This does not show that emotions are not propositional attitudes, but it is enough to cast doubt on the Humean theory.

are understood in terms of reasons, since they suggest that there are widespread differences in understanding human behaviour, emotions and mindedness across cultures. However, it would be too hasty to conclude straightaway that understanding actions in terms of reasons is not universally *present* even if it differs in form in different contexts. There may be different ways that reasons can enter into the understanding of actions, especially if reasons are not to be analysed in terms of beliefs and desires.

Morton suggests that people in other cultures may “make very little use of explicit thoughts about rationality ... and make a lot more use of our capacities simply to do the appropriate or cooperative thing, without producing reasons for it” (2007 p. 214). It should be noted that even if this were true, it would not lead to any kind of scepticism about morality or moral reasons. For one thing, doing what is appropriate and co-operative may often be a way of responding to reasons without thinking in terms of reasons. Moreover, people do not need to talk explicitly about *rationality* in order to talk in terms of reasons and exchange reasons. As Ratcliffe says:

That an action has a reason does not imply that it is the outcome of reasoning. An action performed in response to a situation, without any reasoning, can still be understood and explained in terms of reasons (2007a p. 216)

Responding to reasons is not necessarily a case of explicitly entertaining those reasons and deciding therefore to act in accordance with them. It can, instead, be a matter of knowing immediately what to do through a kind of attunement that does not involve conscious reflection. Furthermore, as Ratcliffe says, “[e]ven when we explain situations in terms of reasons, we often do so by referring to aspects of situations, rather than to psychological states” (2007a p. 186). This is in part because showing something to be reasonable (i.e. that there was a good reason to do it) can be a way of explaining why someone did it. But no doubt it is also because we do not think that for someone to have done something for a particular reason they must have been *thinking of* that reason at the time. Therefore, even if thinking *explicitly* in terms of rationality or reasons for action is not universal, there is no barrier to thinking that acting for reasons is universal. Exchanging reasons can also be thought of as universal. After all saying, ‘don’t go out, it’s raining’ is a way of giving a reason for not going out. It would be very surprising if exchanging reasons, in this sense, were not universal across all cultures.

I think therefore that we should accept Hutto's conclusion that "[t]he evidence, such as it is, should make us cautious of simply assuming that *all* human cultures share an understanding of belief/desire psychology" (2007 p. 116). In fact, we should be cautious of *merely assuming* that *any* human culture utilises belief-desire psychology as their primary method of interpersonal understanding. Such claims need to be substantiated. We should not accept Hutto's equivocation of belief-desire explanation with understanding in terms of reasons. Even if the former turns out not to facilitate our interpersonal understanding, the latter might still have an important role to play.

### 3.9 Commonsense explanations in terms of reason

Ratcliffe says, "[i]f a person walks into a room and sits on a chair, we would not ordinarily describe her psychology in terms of believing or knowing there is a chair and wanting to sit down"<sup>17</sup> (2007a pp.188-9). As Bittner points out, we "practically never give reason explanations in the terms prescribed by this answer" (2001 p. 4). Not only do we sometimes mention just a belief or just a desire; very often we mention neither in explaining action in terms of reasons. Bittner gives the following examples (2001 p. 5):

Jonathan takes out the garbage because it is Wednesday.  
I pulled over because the police ordered me to.  
Reggie opened the window to get a breath of fresh air.  
I took a cab, I could not have made it otherwise.  
David was in high spirits, so I stayed much longer.  
Thank you, I do not eat meat.  
I returned to John. It is not right to leave somebody in the lurch like that.

He rightly points out that these are reason explanations, and that they are perfectly satisfactory despite lacking any reference to any psychological states of the agent. Bittner recognises that the "standard reply here has been again to construe such an explanation as tacitly relying on another one that does refer to states of the agent" (2001 p. 6). The idea that such explanations 'tacitly' rely on explanations in terms of mental states could be interpreted in any of the following ways:

1. Belief-desire explanations can be inferred from non-psychologistic reason explanations, or vice versa.

---

<sup>17</sup> We might assume that she does believe there is a chair there – otherwise she wouldn't sit down. However, this does not mean that commonsense explanations must mention this belief.



2. The true meaning of a non-psychologistic reason explanation is provided by giving a belief-desire explanation.
3. Non-psychologistic reason explanations are really instances of belief-desire explanations.
4. Non-psychologistic reason explanations can only be understood by people because they are (or, are capable of) thinking in terms of belief-desire explanations.
5. Non-psychologistic reason explanations are not genuine explanations, but they seem to explain because they call to mind belief-desire explanations.

None of these claims seem particularly plausible. Furthermore, none of these claims is a plausible way of showing that the Humean theory is commonsense. Consider the following perfectly adequate commonsense reason explanation:

(A) 'I pulled over because the police ordered me to'

Consider the following alternative:

(B) 'I pulled over because I believed that the police were ordering me to and I wanted to do what the police were ordering me to do'

Suppose that the statement (B) is a commonsense explanation. Now consider the following claims:

- 1\* (A) is explanatory because it can be inferred from (B), or vice versa
- 2\* (A) is explanatory because it *means* (B)
- 3\* (A) really is (B)
- 4\* (A) can only be understood because we are capable of understanding (B)
- 5\* (A) does not explain the action of pulling over, but it calls to mind (B) which does

Even if one of these claims is true, it is surely not commonsense. If none of these claims are commonsense, then commonsense allows that (A) is a perfectly adequate explanation independently of any relation to (B). Moreover, (A) is a *more* commonsensical explanation than (B). Humeans might yet try to claim that statements like (A) are only explanatorily adequate because they are really just incomplete Humean

explanations: they explain by pointing us towards Humean explanations. However, since this answer presupposes the Humean theory, it cannot be used until we have some independent reason for accepting the Humean theory. Again, Humeans might claim that explanations might be adequate instances of belief-desire explanation even if one or both of these elements are missing. However, the only way Humeans can maintain this without abandoning their theory is to claim that these kinds of description are only used when those offering and receiving the explanations can be expected to know what desires and beliefs the agent in question had. Again, this presupposes the point at issue. Why should it be that beliefs and desires are required as background knowledge that facilitates explanation? An argument is required to establish this claim. As Bittner says:

[T]he thesis that a desire and a belief are primary reasons, in the sense indicated, does not register an obvious feature of our explanatory practice; it claims that, unobviously, this practice needs to be understood and construed in a certain way. To say this about the thesis is not an objection, to be sure. It is just grounds for requesting an argument for it. (2001 p. 7)

It cannot just be asserted that commonsense explanations must mention both beliefs and desires when the phenomenon that we are trying to account for is the fact that many of them do not. Humeans might insist that for any action, *there is an account available* in terms of the agent's beliefs and desires, even if we do not always give the explanation in such terms. However, as Ratcliffe says:

It will of course be trivially easy to describe most behaviour in terms of propositional attitudes. But it is not clear why a particular behaviour that we do not ordinarily interpret as the outcome of beliefs and desires *should* be re-described in those terms. The mere possibility of doing so does not itself support the move. (2007a p. 99)

It is not just that ordinary descriptions do not always mention beliefs and desires, however. In fact, according to commonsense, it is quite possible, indeed ordinary, for people to do things that they have no desire to do. Platts says, “[w]e perform many intentional actions ... that we apparently do not desire to perform” (1979 p. 256). As Schueler puts it, when we talk about desire in “the more ordinary sense”, it is commonplace that “one can do things one has no desire to do” (1995 p. 1). Schueler gives the following example:

‘I have absolutely no desire whatsoever to go to this meeting,’ I might say to myself as I dutifully trudge off on a sunny Friday afternoon to the last committee meeting of the year<sup>18</sup> (1995 pp. 48-9).

Humeans might claim that such talk is confused. What Schueler means is not that he has *no* desire to go to the meeting, but that although he has a desire to go, he has a stronger desire not to go. There are a number of problems with this response. Firstly, it does not actually facilitate a Humean folk psychological explanation of why he went to the meeting. If the desire not to go were stronger then surely, according to the Humean theory, he wouldn’t go. Humeans might reply that the desire not to go *feels* stronger, but the desire to go is *actually* stronger. But now we have strayed a long way from commonsense, into philosophical territory. Secondly, there doesn’t seem to be any reason to think that such talk is confused, other than that it conflicts with the Humean theory. If the Humean theory is commonsense, it should be commonsense that people never do things that they have no desire to do. True, people are apt to say things like ‘if you did it, you must have wanted to do it’, but then they also say things like ‘I have no desire to do this at all’ or ‘I don’t want to do this at all’, whilst performing the action in question. Commonsense does not appear to straightforwardly support the Humean theory, then.

Again, Humeans may object that whilst people can do things that they have no *intrinsic* desire to do, intrinsic desires are not the only kind of desire. We may do things that we do not intrinsically desire to do, but it can always be shown that they are the means to something else we do desire. This seems *prima facie* plausible, but in order for it to be true, ‘desire’ must be so broadly construed that it includes any way in which an agent might relate to their purpose in pursuing an action. It would seem difficult enough to give a philosophical account that clearly captured this broad notion of desire in such a way that it can still contribute to explanation. Therefore, to assert that such a notion of desire is found in commonsense seems implausible<sup>19</sup>.

Although mentioning beliefs and desires is unnecessary for explanations of action, and people may ordinarily be said to do what they have no desire to do, perhaps

---

<sup>18</sup> This might be a misleading way of saying ‘I expect no pleasure at all to come from attending this meeting’, and might be compatible with me having a desire to go to the meeting to keep on my boss’ good side. This is possible, but (a) there is no reason for insisting that all such cases must be interpreted in this way unless we are already committed to the Humean theory of motivation and (b) this interpretation is less plausible than taking the statement at face value.

these are just exceptions, and actions are generally, for the most part, explained by reference to beliefs and desires. Is this true? Ratcliffe conducted a survey to find out what students thought interacting with and understanding persons consists of. Out of twenty-five responses, the term ‘belief’ appeared only twice, and the term ‘desire’ only once (2007b p. 228). Ratcliffe reports that “the diversity of rather vague responses suggests that there is no explicit, shared, commonsense conception of everyday social understanding” (*ibid.*). When belief-desire psychology was suggested to the students as a mode of interpersonal understanding, it was “not met with any sign of recognition or eager assent” (*ibid.*)<sup>20</sup>. This single study cannot be considered conclusive, but it does seem to indicate that ordinarily people do not think of belief-desire attributions as essential or important in explaining actions.

### 3.10 Are belief-desire pairs sufficient to explain actions?

Is a belief-desire attribution sufficient for giving a commonsense explanation of an action? That is to say, do all appropriate instances of attributing belief-desire pairs provide adequate commonsense explanations by themselves, without the need to mention any additional explanatory factor?<sup>21</sup> This is distinct from the claim that such descriptions are necessary and sufficient for explanation, since it does not assert that they are necessary, and it is left open whether there are other descriptions that are sufficient for explanation. Suppose that we sometimes describe actions in terms of beliefs and desires as proponents of folk psychology would have us believe, are these descriptions, taken on their own, adequate explanations? The answer given herein is that they are not: in no case are they sufficient to explain. Subsequently it is asked whether, when present, they make an important contribution to explanations. Again, it is answered that they do not. In many cases nothing is lost if they are dropped from the explanation.

---

<sup>19</sup> That is to say, once again, the debate has moved into philosophical territory. This is not an objection to the Humean theory; it is an objection to the claim that the Humean theory is commonsense. What is required is a philosophical argument. Such arguments are discussed in chapters four to six.

<sup>20</sup> It should be noted that whilst these students were studying philosophy of mind, they had not received any lectures on folk psychology at all, nor any lectures concerning interpersonal understanding or the explanation of human action.

<sup>21</sup> We should keep in mind the distinction between two kinds of sufficiency made above (fn. 6). Belief-desire pairs are sufficient for explanation in the relevant sense even if mentioning them does not *guarantee* a successful explanation. Of course not just any belief-desire attribution could be sufficient to explain; if such descriptions can explain at all then they can only do so when the belief and desire are attributed in the right way.

The first reason for thinking that belief-desire descriptions are insufficient for explanation is that ‘belief’ and ‘desire’ are vague and ambiguous terms. The term ‘desire’ encompasses a wide variety of states that vary in their role in explanation, and in their phenomenology (Ratcliffe 2007a p.190). Commitments and convictions, for example, play a role in interpersonal understanding, but their nature is obscured by referring to them as ‘desires’ (Ratcliffe 2007a pp. 200-1). The fact that I enjoy cycling may play a role in explaining some of my activities but this cannot usefully and commonsensically be paraphrased as ‘I desire to cycle’. For one thing, there are many ways of enjoying something (Ratcliffe 2007a p. 209) and many things that we enjoy in different ways, such as cups of tea, sex, skydiving, watching football, getting revenge, sleeping in on a Sunday morning or speaking eloquently in a foreign language. Even to refer to these all as kinds of enjoyment is not particularly enlightening; to try to subsume them all under the heading of ‘desire’ is just artificial. For this reason, mentioning that someone ‘desires to cycle’ (and has beliefs about to fulfil this desire) is unlikely to suffice when we are trying to explain why they spend their time as they do.

Earlier we said that it is possible and commonplace for people to do what they have no desire to do. Someone might go to a meeting even though they have no desire to go. There are other cases that Humeans would describe as cases of competing desires: A desires to  $\Phi$ , but also desire to  $\Psi$ , where  $\Psi$ -ing is incompatible with  $\Phi$ -ing. Ratcliffe gives the following example:

Suppose the sun is shining outside, my colleagues are all drinking in the pub around the corner and I would really appreciate a cold beer and a good chat right now. However, I am scheduled to give a lecture on folk psychology during the next hour. I do not relish the prospect of giving the lecture and the beer is much more appealing. But I give the lecture anyway, the reason being that a commitment integral to my professional life over-rides an episodic inclination towards conversation and beer. (2007a p. 200)

The problem with referring to all motivational states as ‘desires’ and distinguishing between only by their differing propositional contents and relative strengths, is that it “obfuscates the difference between them” (*ibid.*). As Ratcliffe points out, “just about any everyday description would serve to convey the nature of the conflict; ‘I’d love a beer right now but I’ve got to give that darned lecture’” (*ibid.*). Humeans might admit that desires are of many phenomenologically diverse kinds, but that all fall under the ordinary concept of ‘desire’. But as a claim about our ordinary explanatory practice, this

is just false; we would not ordinarily explain why Ratcliffe gave the lecture by saying that ‘he wanted to give the lecture’ or ‘he desired to give the lecture’. We would usually say that he gave the lecture *despite* not wanting to, or *even though* he would have preferred a beer. To *insist* that he desired to give the lecture is to impose a philosophical distinction onto our everyday talk rather than to report our ordinary talk.

It is not just that Humeans artificially force phenomenologically diverse states under the label ‘desire’. Insisting that we talk in terms of ‘desires’ or ‘wants’ strips away all the other interesting features of situations that place actions in context and facilitate understanding. Consider the following example given by Frith and Happé (1999 p. 2):

[I]t is trivially easy to explain why John will carry his umbrella with him: it is because he *believes* it will rain and he *wants* to stay dry.

It is no doubt true that if he either thought that it certainly wasn’t going to rain or didn’t care about getting wet, we would lack an explanation of John’s action<sup>22</sup>. However, it does not follow from this that this belief-desire pair constitutes our ordinary explanation of what John does or that they suffice as an enlightening explanation. Indeed, mentioning them is seldom likely to be enlightening. As Ratcliffe points out, this description fails to distinguish differences between scenarios that are conveyed with ease in even in the crudest of everyday narrations, such as whether John took his umbrella with him out of habit or interrupted his routine to fetch it upon observing the dark sky.

It is not just the use of ‘desire’ that fails to distinguish between the many varied contexts in which it can be used; ‘belief’ also displays these features. Morton puts the point in the following way:

Philosophers often write as if we had clear concepts of two relations between a person *a* and a proposition *p* “*a* believes *p*” and “*a* desires that *p*”. In real spoken English we use a great variety of words: thinks, suspects, is of the opinion that...; wants, longs for, would like, has a yen for... And we use contrasts between these words to indicate different kinds of belief and desire. (Morton 2007 p. 216)

---

<sup>22</sup> This does not mean that the action was inexplicable. Many other possible explanations of umbrella carrying spring to mind relating to personal idiosyncrasies, habits, norms, customs, and other uses for umbrellas.

The ambiguity of belief is highlighted by the fact that “a bewildering variety of senses attach to words in foreign languages that have been used as translations for the English verb ‘to believe’” (Kusch 2007 p. 186, cf. Needham 1972). Furthermore, as Kusch says, “the word ‘belief’ has a plethora of different uses in modern English, uses that cannot be reduced to one specific core” (*ibid.*). ‘Believe’ can be used to express uncertainty, ideological commitment, or trust or confidence in a thing, person, cause, course of action or way of life (Ratcliffe 2007a pp. 188-9). The meaning of ‘belief’ is also context dependent. Ratcliffe discusses Needham’s (1972 p. 40) example of someone, upon being asked whether they will accept something, such a biscuit, replying ‘I believe I won’t’:

[I]t is never simply a matter of denying a proposition. It might be a way of politely declining an offer or... an indignant response to an unreasonable request... Other uses of ‘belief’ abound... depending on how it is uttered, ‘I don’t believe it’, can convey astonishment, disappointment or incredulity... Similarly, ‘I can’t believe you did that!’ expresses anger, disgust, embarrassment, disappointment, admiration and so forth. (Ratcliffe 2007a p. 188)

In these cases it is not merely that ‘I don’t believe...’ has an *additional* meaning to that of denying a proposition, but a *different* meaning. In most of these cases the utterance does not commit the speaker to the falsity of the proposition in question. Commonsense alone does not support the idea that ‘belief’ has a single use, indicating the affirmation of a proposition. Even beliefs that do involve affirmation of a proposition can be very different from each other in nature. Ratcliffe considers the example of someone, Jane, trying to get to her best friend’s wedding. Jane’s belief that her train is about to leave is “a state in which experience and action are very closely tied together...infused with emotion” (Ratcliffe 2007a p. 198). This belief is utterly different to her belief that the Eiffel tower is in Paris:

Her running, the sight of the train, the sound of the whistle, the shutting of the doors, a sense of urgency and a background of projects and concerns all blend together seamlessly. Just about any everyday description of the two cases would succeed in communicating the significant differences between them but stating that... actions are caused by ‘beliefs’ and ‘wants’ or ‘desires’ does not. The two ‘beliefs’ seem to be very different states, connected in very different ways to experience and action (2007a p. 198)

This shows not that philosophical uses of ‘belief’ are illegitimate, but that they do not correspond to commonsense (Ratcliffe 2007a p. 187); their use as technical terms must be justified independently of ordinary usage or may legitimately pick out interesting features of ordinary usage. We can conclude therefore that if belief-desire attributions are necessary for action explanations, this can only be a philosophical theory in which ‘belief’ and ‘desire’ function as technical terms. It cannot be commonsense. ‘Belief’ and ‘desire’ accommodate far too many of our concepts to be explanatorily useful<sup>23</sup> (Ratcliffe 2007a p. 187). Rather, they function as “abstract placeholders for a wide range of states that we distinguish with ease in everyday life” (Ratcliffe and Hutto 2007 p. 18).

### 3.11 Norms and shared situations

Since ‘belief’ and ‘desire’ are too vague and ambiguous to facilitate our everyday explanations, descriptions given in these terms will rarely, if ever, contribute to explanation. Perhaps ‘belief’ and ‘desire’ are umbrella terms that classify a whole variety of other mental state terms that we use in explanation in terms of whether they are cognitive, representational states (beliefs) or non-cognitive, motivational states (desires). I claim, however, that our ordinary explanatory practice does not support this view, since it does not support the view that mental state attributions are necessary for action explanations at all.

Dennett (1987) claims that propositional attitude psychology gives us remarkable powers of prediction we would otherwise lack. He claims that without such employing such a psychology ordinary co-operative undertakings, such as a telephone exchange between a man and his wife making dinner arrangements, could not be understood (1987 p. 26). However, Dennett overlooks the fact that a whole host of background assumptions are necessary to facilitate such understanding. For example, Ratcliffe points out, we must know that telephones are used for communicating, that dinners “can, if appropriately orchestrated, be social events... Wine is an appropriate accompaniment to dinner... There are long-term normative commitments between people, which take the form of ‘marriage’” (2007a p. 87). The list of such assumptions is almost endless. Many of them involve social norms, social roles and the functioning

---

<sup>23</sup> To clarify, my claim is that in many contexts mentioning beliefs and desires is not useful since it does not contribute to explanation. In other contexts, mentioned earlier, insisting on describing an agent as having a particular desire also seems false.



of equipment. These assumptions are normative: they are about what is supposed to happen or what ought to be done (Ratcliffe 2007a pp. 87-8). To say that attributing beliefs and desires is the ordinary, commonsensical way of explaining action is therefore highly misleading. Shared assumptions, particularly in the form of the conformance to norms, are doing most of the explanatory work even in cases where beliefs, desires, or both, are actually mentioned. Much of the explanatory work is done by a shared context. Within that shared context, explanations are typically given in terms of norms, roles and person-specific knowledge of character, habit and relationship, including knowledge of factors such as “hobbies, long-term projects, religious and political commitments, phobias, social background, upbringing, political orientation” (Ratcliffe 2007a p. 96).

Against the background of shared norms and situations, pointing out an aspect of the circumstances is often enough to explain an action:

If someone asks why John is carrying an umbrella, ‘it’s raining’ should suffice as an answer, given the assumption that one’s respondent appreciates something of the norms associated with umbrella use. In cases where this does not suffice, the respondent has not appreciated the relevant norms. Reference to John’s beliefs and desires will not help her to do so. (Ratcliffe 2007a p.194)

If Jane does not know that umbrellas are used for keeping dry, mentioning that John believes it is raining and desires to stay dry will not explain to her why he took an umbrella. Even adding that John believes that by taking an umbrella he will stay dry will not help someone who does not understand the norms of umbrella usage<sup>24</sup>. Someone who does not understand that umbrellas are used for keeping dry does not understand why anyone would *think* that taking an umbrella would help to keep them dry. However, if one *does* understand that, there is no need to mention beliefs and desires at all: pointing out that it is raining is enough. Even in cases where behaviour is unconventional, attributing beliefs and desires does not facilitate explanation:

Suppose that everyone else is holding an umbrella while umbrella-less John is leaping around in the rain, singing happily and getting drenched. Would ‘John believes it is raining and he desires to get wet’ suffice as an explanation here? Again, it would not. It does not add anything more to our understanding than ‘he

---

<sup>24</sup> Humeans might respond by asking what understanding the norms of umbrella usage consists in, if not in having certain beliefs. However, even if such understanding were a matter of having certain beliefs, the fact that one needs to have those beliefs in order to understand the action explanation does not mean that those beliefs are themselves *part* of the explanation. This Humean response therefore does nothing to support the claim that actions must have belief-desire explanations.

is leaping around in the rain, grinning and getting increasingly wet'. (Ratcliffe 2007a p.194)

One reason for this is that, as Ratcliffe says, a belief or desire can only contribute to an explanation insofar as it itself is explicable<sup>25</sup>:

If I ask why Fred crossed the road, only to be told that he believed the Loch Ness Monster was chasing him but that it would not cross the road because traffic lights irritated its skin, this would not amount to any kind of reason, given that the mental state contents are far more baffling than the behaviour. A mental state only contributes to a reason for an action if the action is more understandable in the light of that state than it would have been without it. (Ratcliffe 2007a pp.195-6)

Actions can also be explained by pointing to a person's role. As Ratcliffe says, 'she's a fire fighter; it's her job' is an adequate explanation of why someone is running towards a burning house, whilst 'she believes that there is a fire and desires to extinguish it' is not (2007a p. 95). Similarly, expertise can help to explain actions (Ratcliffe 2007a p. 115). In many situations we may only know what to do, and how to do it, because we have relevant expertise. In such cases, one can explain what is going on by mentioning or describing the form of expertise in question, as well as related practices and norms. Attempting to attribute beliefs and desires to the agent would be a quite unnatural way to go about giving an explanation in this context. Humeans might respond by pointing out that theoretical knowledge is often involved in expertise and that its application might involve identifying sub-goals and the means to achieving them. The former could be characterised as desires and the latter as beliefs. However, doing so is a philosophical imposition rather than an ordinary, commonsensical practice. In any case, it is far from clear whether expertise can be adequately accounted for in this way and whether the deployment of theoretical knowledge to identify sub-goals and the means to achieving them is the most important aspect of expertise. Much expertise is embodied, involving the effortless manipulation of complex pieces of equipment. Where expertise does involve theoretical knowledge, this is often knowledge of norms of 'what one does' in situations of this kind, rather than means-end beliefs<sup>26</sup>.

---

<sup>25</sup> This is not to say that it lacks the *logical form* to constitute or contribute to an explanation, but that it is not an illuminating way of describing the action. Explaining something in terms of something that is itself inexplicable is typically not illuminating and tends to add to, rather than reduce, our puzzlement.

### 3.12 Conclusions

Mentioning the beliefs and desires of the agent is neither necessary nor sufficient for commonsense explanations of actions. Commonsense explanations involve pointing to other factors, such as features of the circumstances, habits, emotions, convictions, commitments, character traits, the use of equipment, roles, practices, expertise and norms. Therefore, the Humean theory is not commonsensical and gains no direct support from commonsense. If the Humean theory is true, it must be a philosophical theory employing technical terms and established by philosophical arguments.

---

<sup>26</sup> As Ratcliffe puts it, “expertise involves attuning oneself to a situation in such a way that the situation itself drives and structure’s one’s action” (2007a p. 115).

## Chapter 4. Platts' dilemma

### 4.0 Introduction

We have seen that the Humean theory of motivation is not commonsense and does not draw any support from commonsense either. It might be thought however, that the Humean theory draws support from the philosophy of Hume. I examine this idea and conclude that it does not. Next I examine a dilemma for the Humean theory due to Platts. According to Platts there are two ways of interpreting the Humean theory; one makes the theory vacuous and the other makes it obviously false. The dilemma does not defeat the Humean theory, but it forces Humeans to alter their strategy in ways that ultimately render it unpersuasive. Through explaining this dilemma it becomes clear that Davidson's view in *Actions, Reasons and Causes* (1963), which some (e.g. Smith 1994) take to be Humean, actually represents a permissive anti-Humean view into which the Humean theory is in constant danger of collapse.

### 4.1 Hume and the Humean theory of motivation

Whilst contemporary defenders of the Humean theory of motivation do not generally claim to represent Hume's views, the fact that they choose to identify their view with Hume is surely not arbitrary. Although not all Humeans take their theory to gain support from the writings of Hume, some do at least refer to Hume's arguments and conclusions, and claim Hume as the inspiration of their thoughts. This might in part explain why (as we will see in the next chapter) so few Humeans feel it necessary to explicitly develop arguments for their view. They take the view, or at least the spirit of the view, to have been demonstrated by Hume when he argued that "reason alone can never be a motive to any action of the will" (1739-1740 p. 413). I will now examine whether Hume really did show this, and, if he did, whether it really provides an argument in favour of the Humean theory.

Although the interpretation of Hume's arguments in the *Treatise of Human Nature* remains controversial (see, for example, Baier 1991, Brick 1996, Owen 2000, Millgram 1995, Snare 1995, Stroud 1977), there is a standard way of interpreting Hume's argument, which runs as follows. Hume announces in the *Treatise* that he "shall

endeavour to prove first, that reason alone can never be a motive to any action of the will; and secondly, that it can never oppose passion in the direction of the will.” (p. 413) He later claims, famously, that “Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them” (p. 414). According to Hume, reason, or the understanding, “exerts itself after two different ways, as it judges from demonstration or probability; as it regards the abstract relations of our ideas, or those relations of objects, of which experience only gives us information” (p. 413). In contrast, Hume says, “[a] passion is an original existence, or, if you will, modification of existence, and contains not any representative quality” (p. 415). Hume argues that, “[i]t is impossible, therefore, that this passion can be opposed by, or be contradictory to truth and reason; since this contradiction consists in the disagreement of ideas, considered as copies, with those objects, which they represent” (p. 415). Hume admits of two exceptions to this principle. A passion is contrary to reason if either it is “founded on the supposition of the existence of objects, which really do not exist” (p. 415) or if, “[w]hen in exerting any passion in action, we choose means insufficient for the designed end, and deceive ourselves in our judgment of causes and effects” (p. 415). Hence, “a passion must be accompanied with some false judgment, in order to its being unreasonable; and even then it is not the passion, properly speaking, which is unreasonable, but the judgement” (p. 415). Therefore, Hume famously concludes, “[i]t is not contrary to reason to prefer the destruction of the whole world to the scratching of my finger” (p. 415).

Humeans distinguish between ‘beliefs’ and ‘desires’ rather than ‘reason’ and ‘the passions’. Some Humeans characterise beliefs and desires as mental states with opposite directions of fit to the world: the content of beliefs should be changed to fit the world, whereas the world should be changed to fit with the content of desires.<sup>1</sup> Humeans emphasise that these two kinds of mental state are distinct in that there are no necessary relations between them: for any pair of a belief and a desire, they can be pulled apart modally; it is possible for one to exist without the other. Therefore there is no unitary state that is both a belief and a desire or which is both belief-like and desire-like. Humeans further insist that desires cannot themselves be assessed rationally except insofar as they are based on false beliefs. Note that if Humeans did not insist on this they would allow that reason might guide action through guiding the creation or

---

<sup>1</sup> This idea is attributed to Anscombe (1957) and was subsequently discussed by Platts (1979), used by Smith (1994), and criticised by Zangwill (1998). It is examined in the next chapter.

alteration of the agent's desires – and in this case reason might be the master or author of the passions and not, after all, their slave.<sup>2</sup>

Having set up this distinction, Humeans claim that beliefs alone cannot motivate. Then, echoing Hume's claim that reason's role in action is to choose means for the 'designed' end, Humeans claim that the role of belief in motivation is to direct us to the means of satisfying our desires. There is a simple argument for this claim: a motivated action is an intentional action, and when a person acts intentionally she acts with a purpose. When someone acts with a purpose she acts to bring about something that she wants in some way. To want is to desire. Therefore, all action is performed to satisfy some desire.<sup>3</sup> However, we cannot act to satisfy a desire unless we have some idea of what kind of actions would satisfy that desire. To have such an idea is to have a belief – a belief about how to bring about the end we desire; that is to say, a means-end belief. Therefore we can state the Humean theory as follows:

Whenever an agent performs an intentional action, it must be the case that they have some desire together with a belief that the action would satisfy that desire.

Such belief-desire pairs are motivating reasons for action; motivating reasons for action can function as explanations of actions.

According to most Humeans we must add that explanations of this form are valid on the assumption that the agent in question is rational. The belief and the desire alone are not enough: the agent must put them together in the right way (Smith 1988b; 1994 p. 92). Since motivational or teleological explanation is a form of rational explanation, it fails to apply to cases of irrationality and hence explains only on the assumption that the agent is rational. Irrational agents may have beliefs about how to get what they want but act in ways contrary to those beliefs because of depression, compulsion, weakness of will, or other kinds of irrationality. Hence, the explanation of why an agent acted as she did must mention a desire, a belief about how to satisfy that desire, and, if it is to be complete, the fact that the agent in question was rational.

For now let us set aside the Humean argument to concentrate on whether Hume's argument provides any support for the Humean theory. Hume is generally

---

<sup>2</sup> See, for example, Platts (1988 p. 189)

characterised as an instrumentalist about practical reasoning. However, according to Millgram (1995), Hume was not an instrumentalist but a sceptic. Hume claimed that reason is the slave of the passions. However, according to instrumentalism:

[P]assion is not a reclining pasha who sends reason scurrying off to bring back this or that object of desire; rather, reason returns with further passions, which the initial passions must, on pain of irrationality, adopt (Millgram 1995 p. 76).

Instrumentalism characterises desire not as a slave, but as having a degree of authority over the passions.

Hume presents two arguments for the claim that reason is the slave of the passions. Firstly, he argues that all reasoning is either mathematical or concerns empirical matters of fact. Since neither mathematical reasoning alone, empirical reasoning alone, nor the two together, can produce any practical conclusion, no reasoning can reach a practical conclusion (see Millgram 1995 p. 77). If there is any instrumental reasoning, it must concern empirical matters of fact, and its conclusions must be theoretical, not practical. Its conclusions might be predictions, but not actions, decisions, intentions or imperatives. As Millgram says, the conclusion of Hume's argument is "evidently not that all practical reasoning is instrumental, but that there is no such thing as practical reasoning at all" (1995 p. 77).

Hume's second argument is that since passions do not represent the world, they cannot be true or false, and since reason is concerned only with truth and falsity, passions can neither be approved by, nor contrary to, reason. Therefore, as Millgram says:

Hume is not an instrumentalist. An instrumentalist holds that there is one (but only one) kind of practical reasoning, viz., means-end reasoning. Hume holds the rather more minimalist view that there are *no* legitimate forms of practical reasoning. (1995 p. 78)

Importantly, this means that Hume denies that it is irrational for someone to fail to take the means to the ends that they desire, even though they might be irrational in failing to know what the means to those ends are. Millgram explains the point as follows:

---

<sup>3</sup> We have already seen that this is not commonsense: not only is it not supported by commonsense, it is counter to commonsense. However, as a philosophical argument it has some power. It is argued in the next chapter that there is no interpretation of this argument that renders it successful.

Suppose... I desire a persimmon because I expect it to taste delicious. I, like most people, am built so that, when I realize that the persimmon will not taste as good as I had thought ... I stop wanting the fruit. Because my judgments as to the flavor of persimmons... can be rationally arrived at, and rationally criticized, my desires are sensitive to my reasoning... But this sensitivity is not itself an aspect of rationality, and failure of such sensitivity does not expose one to the criticism that one is being irrational. If I realize that the persimmon is unripe, and continue to desire to eat it, there is no mistake I am making. (1995 p. 79)

Therefore, according to Hume, there is not one, and only one, categorical imperative of rationality, the means-end principle. There are none. Hume's argument appears suspect, however. As Millgram points out, the claim that all reasoning is either empirical or mathematical is "a terrible premise to use in an argument whose conclusion is that there is no such thing as practical reasoning" (1995 p. 80). The premises of Hume's second argument seem similarly to presuppose the point that he is seeking to establish. The claim that reason is only concerned with truth and falsity seems straightforwardly to presuppose that there is no such thing as practical reason, since actions, decisions, intentions and imperatives cannot be true or false. Likewise, the defender of practical reason is unlikely to be convinced by the claim that the passions are not representational.<sup>4</sup> More importantly, there is no reason to think that desires in general may not be rationally assessable. Why shouldn't there be reasons to want some things rather than others? Until some argument is given, distinguishing sharply between reason and passion in this manner is arbitrary and cannot be used to support the substantive claims about these two supposed domains that are being used to reach the sceptical conclusion.

If Hume's arguments beg the question against the defender of practical reason, then why did Hume take them to be valid? The answer is that he thought that there were strong independent reasons, provided by his semantic theory, for thinking that they had to be true. Hume's idea of the passions is not adequately expressed by the contemporary idea of desire. For one thing, "Humean passions differ from the contemporary philosopher's notion of desire in being multitudinous and qualitatively varied" (Millgram 1995 p. 83). Hume thought of passions and emotions such as fear, joy, anger and benevolence as providing us with goals and motives and causing us to act

---

<sup>4</sup> The defender of practical reason might point out, for example, that fear, which can presumably be counted among the passions, usually has an object. Things can be represented as fearful. Fears can both enter into explanations of actions, and are ordinarily thought to be assessable in terms of rationality: some fears are rational and others are not, depending on whether their objects are fearful. The same can be said for many of the other passions.



(Radcliffe 2008). Humeans want to define all these passions in terms of desire, whereas Hume allowed for many distinct kinds of passion. These he distinguished from one another being their “‘peculiar’ feeling” (Millgram 1995 p. 84). Humeans, on the contrary, typically deny that desires must be felt at all. Hence Hume’s account of mental states was radically different from that of contemporary Humeans. Humeans claim that desires are motivational states with propositional contents but Hume, on the basis of his semantic theory, concluded that “mental states either have contents or motivational force, but not both” (*ibid.*). On Hume’s theory, the passions not only do not represent the world, they lack any contents at all; their intentionality is “simulated by causally linking a passion with a content-bearing judgment” (*ibid.*). Reason only manipulates states that have contents, so passions do not enter into reasoning. Thus, Hume’s category of the ‘passions’ is quite unlike the ‘desires’ of the contemporary Humean theory. Since Hume’s assumptions are not accepted by contemporary Humeans, they cannot utilise Hume’s argument. Hume’s argument works only if we accept Hume’s semantic theory; if we reject the claim that the passions lack content then we need to provide a different argument for the claim that they cannot enter into reasoning.

Millgram, having pointed out that Hume’s practical scepticism was forced on him by his semantic theory, a semantic theory that is now thought discredited, writes:

If I am right, instrumentalists err in invoking Hume’s authority not just because they are mistaken in thinking that Hume shares their view, but in that they suppose that Hume’s arguments, perhaps slightly modified, can be adapted to the uses of a contemporary philosopher. They cannot. (1995 p. 87)

This seems true. However, the consequences of Millgram’s argument are not quite what he takes them to be. Millgram takes Smith’s article ‘The Humean Theory of Motivation’ (1987) to be a defence of instrumentalism. However, what Millgram calls ‘instrumentalism’ is the thesis that there is a single valid norm of practical reasoning: that we should will what we believe to be the means to the ends that we desire.<sup>5</sup> This is a thesis about normative reasons for action. However, Smith intends the Humean theory of motivation as a thesis about motivating reasons, about reasons that *explain* action. In this respect, his view about motivation is closer to Hume’s than Millgram thinks. Hume’s argument shows that whilst there is no such thing as practical reasoning, or a normative reason for action, explanations of action will tend to take the form of

---

<sup>5</sup> I consider and reject this thesis in chapter seven.

mentioning, to use the contemporary terminology, a desire and a means-end belief of the agent. And this is what Smith wants to claim.

However, there are a number of serious problems with any attempt to establish a position like Smith's using Hume's arguments. The first we have already mentioned: Hume's theory of semantics is discredited, and his arguments are completely question-begging without it. The second is that if the argument were established, it would establish not just a Humean theory of motivation, but a thorough-going scepticism about practical reason, entailing that there can be no reasons for action and no valid forms of practical deliberation. Very few contemporary metaethicists want to end up with this conclusion; Smith certainly does not, and it is doubtful if many others who call themselves Humeans would either.

The third problem is that these consequences extend to the theory of motivation. Smith maintains that when a person acts, we can describe their action in terms of the reason that motivated them. That reason is given by mentioning (a) that there was something that the agent desired to do and (b) that the agent believed that the action would contribute towards doing just that. This description explains the action on the assumption that the agent was rational. However, if we accept Hume's argument, this cannot be right. A belief-desire pair may explain the action, but they do not constitute a reason, they do not rationalise the action, and no assumption of rationality is necessary to facilitate the explanation. Rather than giving an answer to the question 'why' an action was done, they show that the question has no application<sup>6</sup>.

## 4.2 Platts' dilemma

Platts objected to the Humean theory of motivation on the grounds that it "is either phenomenologically false... or utterly vacuous" (1979 p. 256). Platts' argument posits a dilemma for Humeans but, as we shall see, the argument does not succeed, because it assumes that there are only two possible conceptions of desire and therefore overlooks other important options open to Humeans. Nevertheless, it shows us a great deal about what a successful argument for the Humean theory would have to look like.

---

<sup>6</sup> Radcliffe recognises this, pointing out that "on Hume's view, the beliefs that direct the passions to their goals are not reasons strictly speaking, even though they are necessary to produce the resulting behavior" (2008, np). This sits uncomfortably with Anscombe's suggestion that intentional actions "are the actions to which a certain sense of the question 'Why?' is given application; the sense is of course that in which the answer, if positive, gives a reason for acting." (1957 § 5 p. 9) If we combine this with Hume's scepticism we are left with the conclusion that no actions are intentional.

Platts' argument is based upon a distinction between taking something to be desirable and feeling a desire for that thing. Platts says that:

We perform many intentional actions in [the moral] life that we apparently do not desire to perform. A better description of such cases appears to be that we perform them because we think them desirable.<sup>7</sup> The difficulty of much of moral life then emerges as a consequence of the apparent fact that desiring something and thinking it desirable are both distinct and independent.<sup>8</sup> (1979 p. 256)

Platts' argument, although almost entirely contained in this short quotation, might not be immediately clear. To start with, his objection does not quite connect with the Humean theory of motivation as we have been discussing it. We have been talking in terms of agents having motivating reasons, but Platts puts his point in terms of performing intentional actions. The gap is easily bridged, however, because it is agreed that an action was performed intentionally if and only if the agent had some (motivating) reason for performing it.<sup>9</sup> Therefore, Platts' argument is that there are some cases where thinking desirable, rather than desiring, is the basis for our having a motivating reason for action. Platts is then able to posit a dilemma, saying that the Humean theory can:

[B]e held true by simply claiming that, when acting because we think something desirable, we do indeed desire it. But this is either phenomenologically false, there being nothing in our inner life corresponding to the posited desire, or utterly vacuous, neither content nor motivation being given to the positing of the desire. (1979 p. 256)<sup>10</sup>

<sup>7</sup> I take it that when Platts speaks of thinking it would be desirable for us to act in a particular way he means to include thinking that we are obliged to act in that way, as well as any other way of believing that it would be good or right for us to so act. This does not rule out recognising reasons to regret having to act in that way. Importantly, it is to be contrasted with being inclined or motivated to act in that way.

<sup>8</sup> It doesn't follow that something is desirable from the fact that it is desired (this is the fallacy of which Moore accused Mill; I discuss this issue in chapter seven). Plausibly, however, it might be claimed that considering something desirable involves desiring it at some level. Whether this is right or not, it cannot be of any help to Humeans, since it posits a necessary connection between the belief that  $\Phi$  is desirable and the desire for  $\Phi$ . It is essential to the Humean theory to claim that there are no such necessary connections between beliefs and desires.

<sup>9</sup> Mele says of Davidson that he "is prepared to say, at least, that 'it is (logically) impossible to perform an intentional action without some appropriate reason'" (Mele 1992 p. 6, citing Davidson 1980, p. 264). Note, however, that, according to Mele, there "is no neat and uncontroversial way of sorting... actions into intentional and nonintentional deeds" (1992 p. 6). It is also worth bearing in mind that the distinction between actions and mere events (see e.g. Wittgenstein 1972 § 621; Davidson 1980) is insufficient. As Velleman (2000), following Frankfurt (1988 pp. 58-68), Gardner (1993 pp. 188-9), O'Shaughnessy (1980 ch. 10) and Freud (1960), points out, it is a mistake to think "that human events can be divided without remainder into actions and mere happenings" (Velleman 2000 p. 2). We need at least the three categories of phenomena that Velleman proposes: "mere happenings, mere activities, and actions" (2000 p. 4). However, this does not undermine the claim that those events that fall into the last of these categories (actions) must be intentional, and that this means that they must be done for reasons.

<sup>10</sup> Ratcliffe makes a similar point about folk psychology (FP):

Smith (1994) points out that this seems to assume a phenomenological conception of desire according to which A only desires to  $\Phi$  if A *feels* a desire to  $\Phi$ . However, Smith argues, since the phenomenological conception of desire is false there does not need to be anything in our inner life for it to be true that I desire  $\Phi$ . This blunts the first horn of the dilemma: it can be held that when we act because we think something desirable, we do indeed desire it. However, the second horn of the dilemma awaits. It isn't clear that anything is added to the explanation by positing this desire, and it isn't clear why the desire should be posited. The idea seems to be that when an agent acts intentionally they must in *some sense* have wanted to perform the action. But that was already captured by saying that the agent thought the action desirable. Why insist that they must have had a desire? Why can't the thought that the action was desirable itself feature in the explanation? An argument needs to be given. Smith does give an argument, and we will turn to this in the next chapter. But first let us look more closely at the two horns of this dilemma. Once we understand more clearly why Humeans need to escape from this dilemma we will understand what is involved in arguing for and defending the Humean theory of motivation.

### 4.3 Phenomenological and ordinary conceptions of desire

According to the phenomenological conception of desire, desires are mental states that essentially have phenomenological content. This means that just as one cannot be in pain without knowing that one is in pain, one cannot have a desire without knowing that one has that desire. According to Smith, this is plainly false, and this falsity can be shown by counterexamples of the following kind:

- (a) John desires to look at himself in the mirror while buying his newspaper but sincerely denies that this is so. When the mirrors are removed from the newspaper

---

There are many varieties of motivation, which frequently conflict in all sorts of ways. FP examples tend to pass over such differences and replace them with the empty assumption that the strongest desire will be the one that wins out. If 'desire' is understood in a restrictive sense, this is untrue, as commitments and so forth can defeat desires. And, if a more encompassing sense is employed, it will [be] trivially true, as the 'strongest desire' will be defined as whatever psychological state leads one to act as one does. The latter being the case, appeals to desire will be vacuous when it comes to explaining why one acted as one did. (Ratcliffe 2007a p. 200)

stand, he loses his preference for that particular stand. Therefore, it is possible to have a desire without knowing that one has that desire (see Smith 1994 p. 106).

(b) John believes he genuinely desires to pursue a career in music, but really he only desires to please his mother who is fanatical about him being a great musician. When she dies he loses all interest in music. Therefore, it is possible to believe that one has a desire when one does not (see Smith 1994 p. 106-7)<sup>11</sup>.

According to Smith this is our first reason to reject the phenomenological conception of desire: that it entails, falsely, that we cannot be fallible about our own desires. Therefore, it gets the epistemology of desire wrong. The second reason to reject the phenomenological conception is that it gives us no way to make sense of the fact that desires have propositional content, and no epistemology for the propositional content of desire. This is because the phenomenological conception of desire models desires on sensations, and sensations are states that lack any propositional content. The third reason to reject the phenomenological conception of desire is that it gets the phenomenology of desire wrong: it entails that someone only has a desire when they *feel* that desire. But desires such as “a father’s desire that his children do well” (Smith 1994 p. 109) are thought to persist in moments when they are not actually *felt*. Therefore, the phenomenological conception of desire cannot account for the fact that desires have propositional content, and it gets the epistemology and phenomenology of desire wrong.

This is a relief for Humeans because people often perform actions when they neither *feel* a desire to perform that action nor *feel* a desire for some end that the action is a means to. If the phenomenological account of desire were correct then it would follow that if they *felt* no desire, they *had* no desire. The Humean claim would thereby be shown to be false: people perform actions that do not contribute to the satisfaction of any of their desires. But the phenomenological conception of desire is false, so this objection to the Humean theory cannot get started.

---

<sup>11</sup> John was mistaken because he believed he had an intrinsic, non-derivative desire to pursue a career in music. It turned out that he lacked that desire, but had a non-derivative desire to please his mother and a derivative desire to pursue music that was dependent upon his believing that this would please his mother. In one sense, then, he was not mistaken that he desired to pursue a career in music, but he was mistaken about the nature of this desire. This perhaps shows that this example needs to be clarified in order to support Smith’s point. However, I do not pursue such a clarification since I do not wish to defend Smith’s views on this matter.

Related to the phenomenological conception of desire is a conception of desire discussed by Schueler. Schueler distinguishes “between two senses of the term ‘desire’... the philosophers’ sense, in which... the term ‘desire is understood so broadly as to apply to whatever moves someone to act ...[and] the ordinary sense, in which one can do things one has no desire to do” (1995 p. 1). As Harding says, “We commonly recognize that there are certain things that are good for us, that we should do regardless of whether we really desire to do them” (1999 p. 325). This is the ordinary conception of desire, and is to be contrasted with the ‘philosopher’s sense’ of desire. These two conceptions appear to correspond to the two horns of Platts’ dilemma. The ordinary conception of desire is supposed to reflect our ordinary usage of the concept of desire. Therefore, it cannot be identical with the phenomenological conception of desire. If it were then Smith’s examples – which must be construed as appealing to the ordinary conception of desire – could not show that the phenomenological conception of desire is false.

Schueler also says that when an agent acts intentionally after deliberation, there is a sense in which “the agent must have wanted to do whatever it was he or she did” (1995 p. 2). But there is also a sense in which “it could easily turn out that what moved the agent in the end was not the thought that he or she wanted something or could satisfy some set of desires” (*ibid.*). We should accept this. Our discussion of folk psychology and ordinary explanations of action showed that mentioning desires is neither necessary nor sufficient for explaining action. More importantly, our ordinary conception of desire allows that people may not only do things that they do not want to do, but they may also do things they have no desire to do. Although the phenomenological conception of desire is false, the ordinary conception of desire is not<sup>12</sup>. This means that Platts’ dilemma re-emerges in the form described by Schueler. Ordinarily we say that agents perform actions that they had no desire to perform, for instance because they think that it is desirable that they act in that way or there is a good reason for them to act in that way. This seems to show that the Humean theory is false. Humeans can hold their theory to be true by insisting that when someone thinks something desirable they do indeed desire it. There are two problems with this. Firstly,

---

<sup>12</sup> Note that in one sense the ordinary conception of desire is not something that could be shown to be wrong; it is that against which we determine whether other characterisations of desire are wrong. However, this does not mean it is the only valid conception of desire. ‘Desire’ might be used in a technical sense that particularly brings out one particular aspect of our ordinary usage. The claim that

it is unclear whether this claim is true; some argument needs to be given for it if it is to be the basis for our theory of motivation. Secondly, and more importantly, it seems unlikely that this idea can be of any help to Humeans, since it posits a necessary connection between the belief that  $\Phi$  is desirable and the desire for  $\Phi$ . It is essential to the Humean theory to claim that there are no such necessary connections between beliefs and desires.

Humeans might respond by claiming that thinking  $\Phi$  desirable is not a belief that entails a desire for  $\Phi$ ; rather, thinking  $\Phi$  desirable is *itself* a desire for  $\Phi$ . The same problem then re-emerges, however. If it is claimed that thinking  $\Phi$  desirable is a simply desire and not a belief at all then the position looks implausible and a substantial argument for it is required. If the claim is that thinking  $\Phi$  desirable is both a desire and a belief then the claim is no longer Humean since it posits a necessary connection (identity) between a belief and a desire. Furthermore, this latter option involves departing from the ordinary conception of desire and introducing a broader notion of desire that includes evaluative beliefs. Introducing this technical sense of desire makes the Humean belief-desire claim true *by definition*, *by stipulation*, and hence makes the belief-desire claim vacuous.

Later Schueler expresses the same thought by distinguishing between ‘proper desires’ and ‘pro attitudes’:

A ‘proper’ desire is one where it makes sense to say that someone acted even though she had no desire *at all*, in *this* sense, to do so, as I might attend what I know will be a really boring meeting even though I have no real (i.e. ‘proper’) desire to go. Of course if I attended that meeting of my own free will, even if only because I regarded it as my duty to attend, then there is a *sense* in which I *must* have wanted to go. That sense of ‘want’ is what I am calling the ‘pro attitude’ sense (Schueler 2003 p. 24)<sup>13</sup>

At one point Schueler seems to imply that a proper desire is a “craving, yen, urge” (2003 p. 24). Whether this is so or not, it should not be insisted upon in posing the

---

using ‘desire’ in this way is useful has to be justified of course, and then the danger arises of confusing the technical with the ordinary sense of ‘desire’.

<sup>13</sup> Wallace also recognises this point. He writes:

what is at issue here is not merely whether our intentional actions always involve desires... in a technical sense of the word ‘want’, agents always do what they want to do. This is the sense, commonplace in philosophical discussions, in which it follows from the fact that S did x intentionally that x was something that S wanted to do (1999 p. 237)

dilemma for the Humean theory. The important point is that, as argued in chapter three, in the ordinary sense of desire it is possible to do something that one has no desire to do.

Humeans might insist, however, that whenever we act, we must have a proper desire to do something that would be achieved by what we do, even if we do not have a proper desire to perform the act itself. However, it is not clear why we *must*. It does not seem to be commonsense that we must. More importantly, this presupposes a particular conception of action, the teleological conception of action, according to which actions are always performed in order to bring about some ‘end state’. It is argued in the next chapter that we have no reason to accept this conception of action. If this is right, it would seem not only that we have no reason to believe the Humean claim that action requires a proper desire for an expected outcome of the action. We will also – if there is good reason to reject the teleological conception of action – have a reason to believe that the Humean claim is false.

According to the new version of the dilemma that we have formulated, since the Humean theory is false on the ordinary conception of desire, Humeans are forced to give their theory in terms of the pro attitude conception of desire. Later I will suggest that Humeans do not really have to choose between only these two options. There are other possibilities. But it is important that we understand the pro attitude conception of desire. Then we will see that we should accept that action always involves desire in this sense, and that mentioning pro attitudes is a way to explain actions. However, we will also see that this does not support the Humean theory in any way.

#### 4.4 Pro attitudes and primary reasons

When it comes to the Humean theory of motivation, Davidson’s claims are usually taken as the starting point for the discussion. As Schueler notes, “this is partly because of the brilliant job Davidson did in demolishing the most prominent objections to his two claims” (2003 p. viii). Smith (1987 p. 36; 1988a p. 589; 1994 p. 92) takes himself to be developing and defending the first of Davidson’s (1963 p. 687) claims when he gives the following as the Humean thesis:

- P1      R at t constitutes a motivating reason of agent A to  $\phi$  iff there is some  $\psi$  such that R at t consists of an appropriately related desire of A to  $\psi$  and a belief that were she to  $\phi$  she would  $\psi$ . (Smith 1994 p. 92)



This is a statement of the belief-desire thesis. Identifying a motivating reason in this way identifies what Davidson calls the agent's 'primary reason' (Davidson 1963 p. 686). That primary reasons are identified in this way is the first of Davidson's two claims. Davidson's second claim is that such primary reasons are the causes of actions (*ibid.*). Let us set aside this second claim, and look at whether we have any reason to accept Davidson's first claim in the form he originally presented it. Then we will be in a position to see how Smith's claims differ from Davidson's, and whether we have any reason to accept them.

The first thing to note is that Davidson's concern in 'Actions, Reasons and Causes' was not to argue for P1, but to show that it did not entail that reasons cannot be causes<sup>14</sup>. The second thing to note is that Smith's citation of P1 is not quite faithful to what Davidson actually says. Davidson in fact characterised his thesis about reasons in the following way:

*R* is a primary reason why an agent performed the action *A* under the description *d* only if *R* consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property. (1963 p. 687)

Davidson gives examples of pro attitude, such as "desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions and public and private goals and values" (Davidson 1963 p. 686). We can understand this claim by seeing that it is a way of talking about the purposive nature of intentional actions. Whenever an agent does something intentionally, they do it with some purpose or for some reason. Pointing out what this reason or purpose was involves pointing out the favourable light in which the agent saw their action. Doing this involves pointing out some aspect or feature that the agent took the action to have, given its relation to the circumstances in which it was to be performed. Suppose we call this aspect or feature and its relation to the circumstances, a 'property', and the agent's taking there to be this 'property' a 'belief', we can understand what Davidson means by "a belief of the agent that *A*, under the description *d*, has that property" (1963 p. 687). If mentioning this 'belief' expresses, or partly expresses, the favourable light in which the

---

<sup>14</sup> Whilst Davidson famously defends the view that reasons are causes, Smith claims to remain neutral on this issue. However, this is not the only or the most important difference between Smith's position and Davidson's.

agent saw the action, the purpose for which they performed it, then we can say that the agent has a ‘pro attitude’ towards the ‘property’ mentioned in reporting the ‘belief’. To say that the agent saw an action in a positive light, as worth doing, in virtue of it having a particular property entails that the agent takes it that the presence of this property is a reason in favour of the action. And to say that the agent had a ‘pro attitude’ towards actions of a particular type is just to report this fact. Indeed, we can define ‘pro attitude’ in this way. Davidson puts the point as follows:

We cannot explain why someone did what he did simply by saying the particular action appealed to him; we must indicate what it was about the action that appealed. Whenever someone does something for a reason, therefore, he can be characterized as (a) having some sort of pro attitude towards actions of a certain kind, and (b) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind. (1963 p. 685)

Now, it might seem that we could object to Davidson’s claim, understood in this way, by saying that it is vacuous. After all, it seems we have given an account of reasons in terms of pro attitudes whilst giving an account of pro attitudes that makes it appropriate *by definition* to attribute them whenever an agent acts for a reason. Hence, we have just stipulated that every action is explicable in terms of a belief and a pro attitude. However, that this is so would only seem to be problematic if we expect this account to actually do any explanatory work, or think that it rules out any possible forms of reason-explanation that were previously open to us. I do not think that it does. This account is just another way of putting the point that when we give an agent’s reason for action, their motivating reason, we show the positive light in which they saw the action. That is to say, we point out what they took to be a good reason for performing the action<sup>15</sup>. If we accept this then we should accept the belief / pro attitude account. In mentioning the belief, we mention some aspect of the action that the agent was aware of. In mentioning the pro attitude we point out that (and, possibly, why), given the circumstances, the agent took this aspect to provide a reason<sup>16</sup>.

---

<sup>15</sup> In saying this we do not commit ourselves on the question of whether the agent was right to take what they did to favour acting as they did. We do not even commit ourselves to the possibility that they might have been right – that is, we do not even commit ourselves to the existence of normative reasons.

<sup>16</sup> Davidson is not always understood in this way. Hornsby, for example, says that Davidson’s “model of action explanation links what is done intentionally with the cognitive and affective states of the agent” (1980 p. 79). However, she goes on to say that this shows the agent as seeing the action as a way of bringing about “something that he favours (wants, thinks good, or whatever)” (*ibid.*). This claim presupposes the teleological conception of action so, if the arguments of the next chapter are right, it cannot be true. However, it is important to note that Hornsby confuses two ideas here. In the latter quotation she adopts a pro attitude conception of action explanation. In the former, in talking of

Belief / pro attitude explanations of action are a way of bringing to light the nature of motivating reasons. Motivating reasons are to be distinguished from normative reasons. When there is a normative reason for action, there is some consideration that genuinely favours acting in a particular way. Normative reasons are good reasons. According to Dancy, motivating reasons and normative reasons are not two kinds of reason; the distinction marks two *contexts* in which we use the notion of a reason (2000a p. 2). Giving a motivating reason is a way of answering the question of why an agent acted. It is a way of showing what got them to act, what persuaded them to act, what made them think that acting in that way was a good idea or had something to be said for it. So to give a motivating reason is to mention some consideration that the agent acted in the light of, something that the agent *took* to be a normative reason (see Dancy 2000a pp. 1-6). As Wallace writes:

It is the agent's seeing some consideration as recommending or speaking in favour of her action that is the generic form of a motivating reason, construed as the fact or state of affairs that provides the answer to the explanatory question as to what motivated the agent to act. (2006a p. 68)

In characterising motivating reasons in this way we do not actually commit ourselves to anything about the existence or nature of normative reasons. We only commit ourselves to the view that when people act they take themselves to be acting for reasons. (They might turn out to be mistaken.) This seems undeniable. When someone acts intentionally, she must have thought there was something to be said for acting in that way at the time she decided to do so. And this is what we meant by saying that they had a pro attitude towards some feature and a belief that the action had that feature. We only meant that agents act for what they take to be good reasons, and nothing more.

What of Davidson's claim that *primary* reasons are to be given in this way? There might be many considerations that could potentially have been *the* reason why the agent acted. Of course, there might be more than one consideration that the agent took into account in deciding to do what they did, and there might not be one single reason which, mentioned on its own, constitutes an adequate explanation of why the agent did what they did. Even so, there is a distinction between reasons for which the action *actually* acted and reasons on which the agent *might* have acted but did not. Saying that a reason is a *primary* reason is presumably supposed to indicate this distinction: that the

---

'cognitive and affective states' she seems to have something else in mind – perhaps that actions are to be

reason in question is a reason for which the agent actually acted, rather than, for example, a reason that they were aware of but did not take into account. And it seems right to interpret this, as Davidson would have us, as the idea that the reason (or the having of the reason) in question is the *cause* of the action, so long as we adopt a minimal sense of ‘cause’. Schueler explains this sense of ‘cause’ as follows:

To say that one thing caused another, in this minimal sense of ‘cause’, is really not to *give* a substantive explanation of the one in terms of the other: it is to *claim* that there is such an explanation to be had, and that these events, described in some way or other, will be referred to in that explanation. (2003 p. 16)

This way of understanding reasons, if correct, would not entail that moral judgements cannot motivate without the help of independent desires, even if we were committed to moral cognitivism. On this account, making a moral judgement could be cognitive, and yet also be a pro attitude that might, together with awareness that an action has a particular characteristic, be sufficient to provide a motivating reason. This account does not entail that desires need to be mentioned in explanations of action or that desires need to be present for action. Desires are just one kind of pro attitude that can feature in reason explanations. We need to be careful about terminology, because at this point we are not working with a particular conception of desire. It should be noted that some authors sometimes use the term ‘desire’ in the way that we have been using the term ‘pro attitude’. There is nothing necessarily wrong with doing so, although if we do so we will tend to forget that we have not ruled out that moral judgements and other kind of evaluations may belong to this class, along with the kind of states such as urges, longings and wantings that we would usually more closely associate with the term ‘desire’. Importantly, if we choose to use ‘desire’ in this way we should not think that there is any implication that the class of states it refers to are necessarily non-cognitive, not rationally assessable and not truth-apt. To think that there is any such implication would be to attempt to settle the debate about the nature of motivation by stipulation and to misunderstand both the nature of the debate and what is at stake.

To return to the original dilemma, interpreting the Humean theory according to the pro attitude conception of desire makes it true. However, interpreted this way, it is hardly the claim that Humeans seem to want to advance. The pro attitude conception of action allows that it is a sufficient explanation of an action to point out something that

---

explained by beliefs in combination with desires in the ordinary sense.

the agent took to favour acting in the way in question. On this conception of action, explanations of action may just involve pointing out features of the action or circumstances that the agent was aware of, or may involve mentioning the agent's beliefs or moral commitments. It need not involve mentioning anything that we would ordinarily call a desire.

This shows us that we need a good argument if we are going to accept the Humean theory. Schueler (2003 pp. 22-4) discusses Blackburn's (1998) claim that no belief can explain an agent's action unless the agent cares about the object of that belief (Blackburn 1998 p. 90). Schueler points out that this is not an argument for the Humean theory, since it "employs the claim that proper desires are always required for the explanation of action *as an assumption*" (2003 pp. 23-4). However, what Schueler says is only true if caring is a species of proper desire. It is far from obvious that this is the case. Frankfurt writes:

It is quite common for people to want various things without actually caring about them, and to prefer satisfying one of their desires rather than another without regarding the object of either desire as being of any importance to them... If satisfying a desire were the only way to avoid being frustrated, then whoever cared about avoiding frustration would have to care about getting the things he wants. However, people can also avoid frustration without getting the objects of their desires by ceasing to want them. (1999 pp. 157-8)

Furthermore, for all we have seen, caring, or even desires, might be rationally assessable. So Blackburn's argument shows even less than Schueler thinks. We can agree with Schueler, therefore, when he writes:

[Blackburn's argument] is completely consistent with genuine beliefs motivating actions *all by themselves* if we distinguish proper desires, such as cravings and urges, from 'pro attitudes', understood simply as '[any] mental states that move us', and understand 'caring' about something in terms of the latter. (2003 p. 24; cf. Schueler 1995 pp. 29-38)

A belief that something is my duty might be a pro attitude in this sense; in this sense if one does one's duty then one must *want* to do one's duty. However, as Schueler points out, "it doesn't follow that this pro attitude *consisted of* anything other than my belief that I had a contractual duty to do it" (2003 p. 24). That is, I must have wanted to do my duty, but "it doesn't follow *at all* from this that what moved me was some *independent* mental state, a proper desire... to do what was morally right" (2003 p. 24). That the

agent has a pro attitude does not show, without additional argument, that the agent had a desire in *any* other sense. It certainly does not show that they have a proper desire or a desire in the ordinary sense. The Humean theory is not supported in any way by the discovery that pro attitudes are required for explanations of action. Indeed, this is not a discovery at all. It is true by definition. What is needed, then, is an argument for the Humean theory.

## 4.5 A constitutive explanation of action

Smith claims that the Humean theory of motivation gives a constitutive answer to the question of what explains an agent's actions. He claims:

actions are to be explained by an agent's wanting some outcome and his believing that he is doing something that he can just do that will bring that wanted outcome about... If we have an event for which such a belief-desire explanation cannot be given then, according to Humeanism, we do not have an action at all, but rather something that merely happened. (2003 p. 460)

To see why Smith's claims about motivating reasons entail this, we need only note that intentional actions are the only kind of actions there are. An unintentional action is not an action at all, but "rather something that merely happened" (Smith 2003 p. 460)<sup>17</sup>. Unintentional 'actions' are not done for reasons, but all intentional actions are. Since all actions are intentional, and all intentional actions are done for (motivating) reasons, all actions are done for reasons; being done for a reason is constitutive of action. If a motivating reason is an appropriately related means-end-belief-desire pair then for every action the agent who acted must have had such a belief-desire pair. Moreover, since being done for a reason is constitutive of action, mentioning the reason for action is the only way to explain the action *qua* action. Hence, the Humean theory is a constitutive answer to the question of what explains an agent's action.<sup>18</sup>

---

<sup>17</sup> Velleman's distinction applies again here: if something is not an action, it might not be 'something that merely happened' – it might be a 'mere activity' but not a mere event (see Velleman 2000 pp. 2-4). So there might not be Humean explanations of everything we do – they might not always be available for mere activities since the Humean constitutive claim is restricted to actions. Unexpected consequences of our actions, or things that we do carelessly (think of a careless driver who knocks down a pedestrian), are also neither intentional, nor merely things that happen. They are things that we are involved in bringing about, but that we do not set out to cause. They are not strictly actions of ours, nor are they mere events; they are not activities either. Since my concern, particularly later in the thesis, is with what we find ourselves to have reason to do when we reflectively deliberate, these acts that fall short of deliberate or intentional action are not problematic for my account.

<sup>18</sup> Nowell-Smith suggests this view when he writes: "A man can choose without having any reason for his choice in sheer absence of mind or from sheer force of habit. These are minimal cases that hardly deserve

Given that the Humean theory is a constitutive explanation of action, the worry now emerges that it may, after all, be vacuous. This worry arises because the Humean theory is dependent upon a dispositional conception of desire. To see why this is a worry, consider Schueler's distinction between two senses of desire. Firstly, he says, there is:

the philosophers' sense, in which, as G. E. M. Anscombe says "the primitive sign of wanting is trying to get," that is, the sense in which desires are so to speak automatically tied to actions because the term 'desire' is understood so broadly as to apply to whatever moves someone to act. (Schueler 1995 p. 1; c.f. Anscombe 1957, § 36)

Secondly, there is desire in "the more ordinary sense, in which one can do things one has no desire to do" (Schueler 1995 p. 1). The second of these options is not available to Smith's theory, for the following reason. Every action is done for a motivating reason and, according to Smith, "having a motivating reason just is, *inter alia*, having a goal" (1994 p. 116) and "having a goal just is desiring" (1994 p. 116). Therefore, if action entails having a motivating reason, it also entails having a desire. So Smith does not accept that one can do things one has no desire to do.

Smith is left, therefore, with the other sense of desire: desire is to be understood as to apply to whatever moves someone to act. However, if the Humean theory is a constitutive explanation of action then it must provide criteria for distinguishing between actions and non-actions. But if desire is understood as anything that could move an agent to *act* then it presupposes the notion of action and cannot provide a criterion for distinguishing between actions and non-actions. It might seem that we are being unfair to Smith. The two conceptions of desire given by Schueler are not exhaustive, and do not include Smith's preferred option: the dispositional conception of desire. According to the dispositional conception of desire "we should think of desiring to  $\Phi$  as having a certain set of dispositions, the disposition to  $\Psi$  in conditions C, the disposition to  $\chi$  in conditions C', and so on" (Smith 1994 p. 113). This does not mention action, for it is not specified that  $\Psi$  and  $\chi$  are actions. Therefore, it seems that the dispositional conception of desire allows Smith to escape the charge of circularity.

---

the name of 'choice', and I shall consider only cases where a man has reasons for his choice." (1954 p. 102) He also says "To say, for example, that every action must have a motive is to state a tautology, since what a man 'did' without a motive would not count as an 'action'" (1954 p. 114).

Suppose we grant the validity of this response. A constitutive explanation of action is supposed to tell us whether something is an (intentional, reasoned) action because it is supposed to explain actions by picking out that which distinguishes them from everything else. Therefore, Humean explanations should enable us to pick out actions from among the other events and should not depend upon our already being able to do so. We can ask: why shouldn't any event whatsoever count as a candidate for an action, and therefore count as a potential situation in which we could ascribe a desire? In the case of animal behaviour, or the 'behaviour' of plants when, for example, they grow towards the light, Humeans answer that being capable of action and desires requires being capable of belief, and animals and plants are not capable of having beliefs.

We can also ask Humeans why it is that tripping over, for example, is not (normally) an action. Ordinarily we would say that it is not an action because people do not (normally) trip over intentionally, on purpose. Ordinarily people do not trip over for a reason. Humeans need to deliver this conclusion without presupposing it, since they claim that whether a belief-desire pair explains the event determines whether or not it was an action. It is not enough for Humeans just to point out which dispositions and means-end beliefs the agent has. Someone might believe that hitting his foot against the curb would make him trip up, and that tripping up would make the girl notice him, and he might be disposed towards doing things that make the girl notice him. But this still does not tell us whether or not he tripped up on purpose, whether his tripping up was an action. We need to know whether or not the relevant beliefs and dispositions explain the action, whether the agent acted upon that reason. It seems, therefore, that we need to know what caused the agent to trip. We might use cause in Schueler's minimal sense here. We want to know: would he have tripped if he hadn't believed that, were he to trip, the girl would notice him? The causing also needs to happen in the right way: if he tripped because the girl made him nervous then we will not think he did it intentionally. But the Humean theory does not tell us anything about this.

Velleman discusses a case of 'mere activity' reported by Freud for which a Humean explanation can be given. Freud reports breaking his inkstand in an apparent act of clumsiness. However, Freud suggests, this act of apparent clumsiness might be interpreted in light of his sister's earlier remark: "Now the desk really looks very well, only the inkstand does not match. You must get a prettier one." Freud asks: "Did I perhaps conclude from my sister's words that she intended to present me with a prettier



inkstand on the next festive occasion, and did I shatter the unsightly old one in order to force her to carry out her signified intention?” (Freud 1914 pp. 185-6)<sup>19</sup> This is not a case of intentional action – it is a case of clumsy unintentional action. Yet it seems to admit of a Humean explanation. Freud desired to force his sister to carry out her intention, and believed that breaking the inkstand was a way to do just that. Together this belief and desire explain the action, and if either were not present the event would not admit of this kind of explanation and could not be considered an action. However, the event in question is not an action in the ordinary straightforward sense. It does not relate to agency in the way that most actions do. It is unclear whether we should say that the agent really meant to act as he did. On the whole, it seems that we should say that he did not and that the action was unintentional. The agent was not even *aware* of having performed the act in question until after it had happened and was not ever aware of having *decided* to perform the action at all<sup>20</sup>. Attributing a belief-desire pair to the agent or positing their causal involvement does not enable us to understand the event as the kind of action that it is.

The Humean theory, as it stands, is quite inadequate as a theory of action. Just saying that the belief-desire pairs explain actions is not enough. Saying that they explain actions ‘when they relate to them in the right kind of way’ is a useless cop out. Whatever this ‘kind of way’ is, it is this which makes the difference between something that the agent purposely *does* and something that happens *to the agent*, in spite of the agent’s involvement. It is the difference between these two kinds of event, and variations thereof, that we are interested in when we attempt to give a theory of action. If the Humean theory is to be plausible, explaining this difference must be central to their account, and they must be careful that they do not end up accounting for things that merely accompany actions, rather than accounting for the nature of action itself. Saying this certainly does not constitute a refutation of the Humean theory, but it does indicate that, as it stands, the Humean theory is one that we have no reason to accept. It fails to distinguish adequately between actions and mere events or to describe that which is distinctive about action.

---

<sup>19</sup> See Velleman 2000 pp. 2-3. Velleman cites the translation of Freud’s ‘Psychopathology of Everyday Life’ printed in *The Standard Edition of the Complete Psychological Works of Sigmund Freud* (1960), whereas I quote Freud’s *Psychopathology of Everyday Life* (1914).

## 4.6 Conclusions

The Humean theory of motivation can draw no support from the philosophy of Hume. If Humeans try to use Hume's arguments they will end up with a thorough-going scepticism about practical reason, and a theory of motivation that ignores what is distinctive about intentional action or acting for a reason. However, there is little danger that Humeans will end up in this position, since Hume's arguments, whilst valid, are unsound, beginning from premises that are widely discredited and which no contemporary Humean wishes to accept. If Humeans cannot draw on the legacy of Hume and find no direct support for their theory from commonsense, they need to provide us with some strong independent reasons for accepting their theory. However, in so doing they will need to negotiate Platts' dilemma: the Humean theory is in danger of becoming either utterly implausible, completely vacuous, or of collapsing into a permissive anti-Humeanism. Furthermore, we have an initial reason for thinking that, even if there is a good argument for the Humean theory, it is not a particularly satisfactory theory as it stands, since it fails to capture what is distinctive about action and agency. Nonetheless, we need to consider whether or not there are any good arguments for the Humean theory. If there are then we may need to work on producing a modified version of the theory; if there are not then we may happily reject it. In the next two chapters, therefore, I consider whether we have any reason to accept the Humean theory of motivation.

---

<sup>20</sup> Many of Freud's examples are of this kind, such as those relating to the forgetting of names (1914 pp. 1-14, 27-54), in which the explanation of the incident is in terms of a desire that the agent is not consciously aware of and the incident itself is not an intentional action or not even an action at all.

## Chapter 5. Smith's argument for the Humean theory

### 5.0 Introduction

In chapter four it was noted that whilst Smith takes himself to be following Davidson in defending the Humean theory, his account differs from Davidson's in claiming that motivating reasons are constituted by beliefs and desires rather than by beliefs and pro attitudes<sup>1</sup>. I suggested that Smith's claim, unlike Davidson's, is a substantive claim about what motivating reasons are and that therefore Smith, unlike Davidson, needs to provide a substantial argument for his position. In this chapter I examine the argument that Smith gives. In section 5.2, I question whether its first premise is based on a false assumption, the teleological conception of action. It is concluded that whilst the teleological conception of action is false, the premise can be reinterpreted to avoid presupposing this thesis. I then focus on how Smith's argument uses the metaphor of 'direction of fit' to characterise beliefs and desires and relate them to action. I argue in sections 5.3 and 5.4 that the Humean argument is based on implausible views of belief and the relationship between belief and action. In section 5.5 two objections to Smith's characterisation of desire are discussed and found inconclusive. Section 5.6 examines Smith's claim that desires, as dispositions, can be analysed in terms of conditionals. It is concluded that this claim is unsustainable. In sections 5.7 to 5.9 other objections to the dispositional conception of desire are advanced. It is argued that the dispositional conception of desire is both intrinsically unacceptable and unacceptable when combined with the Humean theory. Hence, it is concluded in section 5.10 that Smith's argument is wholly unsuccessful. It does not succeed in establishing even the slightest advantage for the Humean theory.

### 5.1 Smith's argument

The Humean theory of motivation is the following claim:

---

<sup>1</sup> Smith also, unlike Davidson, claims to remain neutral on whether reasons are causes. Another important difference is that Smith explicitly states that belief-desire descriptions of actions only explain on the assumption that the agent is practically rational. This is discussed in chapter six.

P1 R at t constitutes a motivating reason of agent A to  $\Phi$  iff there is some  $\psi$  such that R at t consists of an appropriately related desire of A to  $\psi$  and a belief that were she to  $\Phi$  she would  $\psi$ . (Smith 1994 p. 92)

Smith writes:

[T]he Humean believes P1 because P1 is entailed by the following three premises:

- (a) Having a motivating reason is, *inter alia*, having a goal
- (b) Having a goal is being in a state with which the world must fit
- (c) Being in a state with which the world must fit is desiring (Smith 1994 p. 116)

Smith's argument is supposed to entail P1 because the first two premises together entail that having a motivating reason is, *inter alia*, being in a state with which the world must fit. This, taken together with the third premise entails that:

- (d) Having a motivating reason is, *inter alia*, desiring.

This supposedly secures the Humean conclusion. To see why, consider again Davidson's view. In order to give the agent's primary reason we have to mention not only a pro attitude, but also a belief that the action instantiated the property towards which the agent had the pro attitude. If we consider the Humean theory as a variant of that claim, substituting 'desire' for 'pro attitude', we can see why it is thought that, once we have identified a desire, a belief, and only a belief, is necessary to complete the reason explanation.

Schueler (2003 p. 26) points out that Smith's argument is not meant merely as a way of asserting that actions are goal-directed, but as an account of goal-directedness in terms of two distinct kinds of psychological state. Clearly the intended conclusion is *not* that reasons explanations of actions involve mentioning beliefs and pro attitudes. This is true by definition, not requiring an argument involving introducing philosophical terminology such as 'direction of fit'. The conclusion, P1, is supposed to involve a much more restrictive conception of desire than the pro attitude conception. The argument is supposed to show that reasons explanations must be understood in terms of

mental states defined in terms of their ‘direction of fit’, where there are exactly two directions of fit, and exactly one state with each direction of fit is required for an explanation. Therefore, much rests upon the argument for the claim that beliefs and desires should be understood as states with different directions of fit. Importantly, this involves a number of claims:

- (a) the notion of ‘direction of fit’, and its two kinds, ‘world-to-mind’ and ‘mind-to-world’, are an accurate way of characterising the nature of some (or all) of our mental states
- (b) no mental state can have both ‘world-to-mind’ *and* ‘mind-to-world’ direction of fit
- (c) having a goal is to be understood as having a mental state with a ‘world-to-mind’ direction of fit
- (d) acting with a goal is to be understood as having a mental state with a ‘world-to-mind’ direction of fit along with a state with a ‘mind-to-world’ direction of fit

From this we reach the Humean theory by adding that the states in question are beliefs and desires, and that the belief and desire are appropriately related to each other and to the action<sup>2</sup>. However, we must also add the first premise of Smith’s argument: whenever one acts (or acts intentionally, for a reason) one’s action is explained by one’s having a goal.

## 5.2 The teleological conception of action

Is the first premise of Smith’s argument true: are intentional actions necessarily goal directed? In one sense they must be. Saying that actions are goal directed is just another way of affirming their intentional or purposive nature. To say that in action an agent has a goal is just to say that they act for a reason, for a purpose, with an intention. No one decides to act unless they see their action in some favourable light. This is not disputed. However, there is another way of interpreting the claim that intentional actions are goal directed. Having a goal can be understood as “intending to bring about some state of affairs” (Shafer-Landau 2003 p. 135), some ‘end state’. Let us call the claim that all action is goal directed in this sense the ‘teleological conception of action’.

---

<sup>2</sup> We must also add the assumption that the agent is means-end rational. This will be discussed in chapter six but for now I wish to pass over any complications that this might introduce.

Interpreted this way, the claim that having a motivating reason is having a goal is false. As Shafer-Landau (2003 p. 135) says, when one acts from duty or because the action itself is worth doing, one's reason is not constituted by the aim of bringing about a state of affairs or 'end state'. Since actions from duty or intrinsic worth are done intentionally, intentional actions are not necessarily done in order to bring about some state of affairs or 'end state'. This is important. If actions necessarily aim at realising 'end states' this suggests motivational explanations must take means-end form. The idea that motivation involves wanting to realise an 'end state' and having an idea about how to do so might then seem quite natural, making the Humean theory immediately attractive. However, seeing actions as necessarily attempts to produce desired 'end states' actually presupposes much of what Humeans affirm and anti-Humeans deny. Anti-Humeans claim that reasons for action can take a variety of forms of which prospective 'end states' are only one. The teleological conception of action is therefore is not only false but begs the question against the anti-Humean.

Importantly, in rejecting the teleological conception of action we take no stance on whether morality is teleologically or deontologically structured; we do not presuppose the falsity of consequentialism. As Shafer-Landau says, "[t]he important issue here is not whether morality actually embodies a teleological structure, but rather whether agents think it does" (2003 p. 135). Consider the following claims:

- (a) people ought always to decide how to act based on the believed consequences
- (b) people always decide how to act based on the believed consequences

Neither entails the other. Even if morality is thoroughly consequentialist, practical reasoning is not. People can perform actions for reasons that are not partially constituted by an end state they are aiming to produce<sup>3</sup>. Therefore, the teleological conception of action is false. These considerations indicate that Humeans are not entitled to use the teleological conception of action as a premise in their argument since it is question begging, unsupported, and false.

---

<sup>3</sup> Pollard (2006 pp. 229-30) also makes this point, but in the cases he discusses it is unclear whether the agent acted for a reason. He mentions actions rationalised by character traits and social conventions. Since these fall within the rational domain but outside the domain of the teleological conception of action they appear problematic for Humeans. However, I only mention them in passing since it is unclear whether they are relevant to my main concern: the nature of motivating reasons.

### 5.3 Beliefs about action

Even if the teleological conception of action is false, the belief-desire thesis might be true. The belief-desire thesis is the claim that a belief-desire pair is necessary to explain any action. A belief-desire pair is given when a desire is mentioned together with a belief about how to satisfy that desire. However, not all desires are desires for states of affairs or ‘end states’. One can desire to perform a particular kind of action independently of what it brings about, and one can believe that a particular action is of that kind. Such beliefs are not instrumental: although they are about how to do what one wants, they are not about the means to realise a desired state of affairs. Let us refer collectively to both these kinds of belief, and to any other belief that might feature in a belief-desire pair, as ‘beliefs about action’. ‘Beliefs about action’ include any belief predicating anything of an action that might possibly form the propositional content of a desire. Since the category of ‘beliefs about action’ is broader than the category of instrumental beliefs, the belief-desire thesis might be true even though the teleological conception of action is false. However, I argue that since one need not even have a ‘belief about action’ in order to act, the belief-desire thesis is false.

There is a simple reason for thinking that action *does* require ‘belief about action’. If one acts intentionally then one has some goal or purpose; there is something that one intends, even if it is not an end state. If one is really to  $\Phi$  or even really to intend to  $\Phi$  (rather than merely to wish that one could  $\Phi$ ) then one must have some idea of how to  $\Phi$ . If one has some idea of how to  $\Phi$  then one has a belief that something one can do,  $\Psi$ , is a way to  $\Phi$ , and this is exactly what the belief-desire thesis says. However, the appeal of this simple argument is lost once we see that ‘having some idea’ admits of two interpretations. Under one interpretation, ‘having some idea of’ means ‘having some belief about’. In this sense it entails that action involves ‘beliefs about action’. Under the other interpretation, ‘having some idea of’ means knowing-how.

Knowing-how is distinct from and not reducible to knowing-that (Ryle 1946; 1949 ch. 2). In saying this we do not have to take a stand on the controversial question of whether knowing-how can be adequately characterised in terms of having an ability (see Fantl 2008). These authors affirm an ‘intellectualism’ that Ryle rejects. Ryle offers a regress argument for the claim that knowing-how and knowing-that are distinct. If they are distinct then we can think of knowing-how as facilitating intelligent action.

However, if knowing-how is a form of knowing-that then intelligent action must be facilitated by contemplation of propositional knowledge. However, Ryle claims:

The consideration of propositions is itself an operation the execution of which can be more or less intelligent, less or more stupid. But if, for any operation to be intelligently executed, a prior theoretical operation had first to be performed and performed intelligently, it would be a logical impossibility for anyone ever to break into the circle (1949 p. 30)

Ginet (1975) criticises Ryle's regress argument, claiming that intelligent action might be performed directly on the basis of knowing-that without the need for an act of mental contemplation. However, Ryle has another, stronger argument. Two people might differ in what they know how to do despite knowing all the same propositions. Ryle's example is of a good chess player who tells a bad player all the rules and tactics that he knows, imparting all of his propositional knowledge about chess. However, the bad player may continue to play badly because he cannot apply this knowledge, he does not know how (Ryle 1949 pp. 215-7 cf. Fantl 2008 p. 454). The intellectualist might reply that more knowledge-that is required: e.g. knowledge that rule *r* applies when the pieces are in position *p*. They may claim that it is only because knowledge-that of so *many* propositions is required that we are tempted to say that knowledge-that is insufficient. This misses the point. In the original example the good chess player, by hypothesis, also lacked this additional knowledge-that, but possessing know-how as he did, he is able to judge and respond appropriately when position *p* arises. There are two further points that we can draw from the example. Firstly, when the good chess player imparts their knowledge it is likely that they may have to reflect in order to formulate their know-how into propositions. In this respect, they may be like a tennis player who knows how to play a backhand return, but who needs to imagine or actually act out playing the shot, and then reflect, in order to be able to state in propositions what bodily movements are involved. Secondly, both the chess player and tennis player may learn how to do what they do without ever having knowledge of the propositions that express the way they do so<sup>4</sup>. In claiming that knowing-how is not reducible to knowing-that we do not commit ourselves to Ryle's positive account of knowing-how, that know-how should be understood in terms of dispositions or as reducible to abilities.

---

<sup>4</sup> This would not seem to be the typical situation – a good tennis player typically knows that 'this is the way to do it', but the fact that it is *possible* to know how without knowing that shows that the former cannot be reduced to the latter.



Stanley and Williamson (2001) propose the following reductive analysis of knowing-how: “S knows how to  $\Phi$  just in case there is a way,  $w$ , such that S knows that  $w$  is a way to  $\Phi$ ” (p. 425). However someone may know that  $w$  is a way to  $\Phi$  without knowing how to  $\Phi$  – someone may know that moving one’s legs in motions  $m$  is a way to cycle, yet fall off every time they try to ride a bicycle. Therefore, Stanley and Williamson add that it is also necessary that  $w$  is presented to S in a practical mode. This is problematic. It deals with the kind of counterexample just indicated only if  $w$  being presented in a practical mode necessarily involves instantiating  $w$ . Fantl writes:

[W]hen it comes to ways I might ride a bicycle, it looks like, in order to be presented with those ways under a practical mode, I will have to know how to instantiate those ways (2008 p. 461).

The reductive analysis fails, since it presupposes the notion of knowing-how that it seeks to reduce (Fantl 2008 p. 460; Koethe 2002 p. 327). Fantl suggests two simple reasons for thinking that knowing-how and knowing-that are distinct (2008 p. 462):

1. Knowing how to do something involves being essentially related to only an act. Knowing that something is the case involves being essentially related to only a proposition. Therefore, knowing-how is distinct from knowing-that.
2. Knowing how to do something is essentially having an ability. Knowing that something is the case is essentially having a belief. To have a belief is not essentially to have an ability. Therefore, knowing-how is distinct from knowing-that.

Nevertheless, Fantl suggests that the important question is whether knowing-how and knowing-that are *normatively* distinct, because it is being normatively distinct that makes them distinct kinds of *knowing*. Fantl writes:

Knowing how to argue is not just the ability to draw conclusions from premises, but to do so according to valid inference rules. Knowing-how and knowing-that, then, are of a normative kind (2008 p. 463)

Suppose that knowing-how and knowing-that are both assessable by norms and that both kinds of knowledge claim must be justified. This does not entail that one is reducible to the other. Suppose that knowing how to play tennis involves not only being

able, during a particular match, to produce good serves, returns, volleys, passing shots, and so on, but also having mastered techniques that reliably produce these results. It therefore involves conforming to norms that determine what it is to play well. It does not follow that this ability is facilitated by beliefs about which body movements produce good shots. The tennis player may know how to play without being able to express this in propositional form. They may have learnt through watching and imitating excellent tennis players or by following their coach's instructions. Importantly for our purposes, there is nothing to suggest that knowing how to  $\Phi$  must be facilitated by, explained by, or even involve having, any belief that  $w$  is a way to  $\Phi$ .

Knowing-how is not propositional knowledge. Very often it is immediate in the form of embodied skills involving complex sets of co-ordinated and sequential movements; in some cases agents may lack beliefs about these movements and might even be unable to adequately describe them in words. Hornsby gives the example of typists who can tell us after only momentary reflection the movement needed to type a particular letter. She notes that this might convince us that typists have beliefs about these movements. However, she continues:

[W]e can still wonder whether it is *because* they have beliefs of this sort that they make the movements they do when they type. It might be rather that they make these movements only because they are the movements that need to be made to type and they know how to type (1980 p. 82)

Other examples suggest that agents might even lack such beliefs. Ruben (1997 p. 236) discusses the movements involved in swimming. Although particular arm movements constituent a way to swim, one need not have any beliefs about these movements in order to swim. Ruben writes:

A person may just know how to do something, having watched and emulated, without being able to say what he does, or without even having available the concepts with which to express this information. (1997 p. 238)

Beliefs about action appear unnecessary: one may know how to  $\Phi$  without having any belief that  $w$  is a way to  $\Phi$ . Furthermore, the idea that beliefs about action are necessary seems intrinsically problematic. Suppose I have decided to  $\Phi$  but suppose that in order to  $\Phi$  I must have a belief about some act,  $\Psi$ , that by  $\Psi$ -ing I would  $\Phi$ . I decide to  $\Psi$  so that I can  $\Phi$ . But now I must have some idea of how to  $\Psi$ . So I must believe that  $\chi$ -ing is a way to  $\Psi$ . An infinite regress threatens. There is a simple and obvious way to block

this regress: deny that knowing how to  $\Phi$  must be understood in terms of having a belief about a way to  $\Phi$ . There must be an irreducible role for knowing-how. Agents can perform actions without having ‘beliefs about action’. The belief-desire thesis is false.

## 5.4 Belief and ‘direction of fit’

We have rejected the teleological conception of action and the idea that action requires ‘beliefs about action’. However, the Humean theory is not yet defeated. Perhaps even when agents lack ‘beliefs about action’, ascribing such beliefs to them is necessary to make sense of their action as goal-directed. Furthermore, ‘beliefs about action’ play a relatively small role in Smith’s argument, only being brought in after the main point has been established – that having a motivating reason is, *inter alia*, desiring. Although we have rejected the belief-desire thesis, we have not undermined the main Humean claims about the relation between action, desire and rationality. It might still be true that action necessarily has *contingent* desire at its source, and that rational norms cannot motivate independently of prior desires to conform to them. Therefore we must examine Smith’s argument for the claim that having a motivating reason is, *inter alia*, desiring. Recall the second and third premises of Smith’s argument:

- (b) Having a goal is being in a state with which the world must fit
- (c) Being in a state with which the world must fit is desiring (1994 p. 116)

Smith claims that ‘having a goal’ and ‘desiring’ can both be characterised by their ‘direction of fit’ as ‘states with which the world must fit’. This is supported by claiming that mental states should be classified according to two ways of relating to the world which are the same in *kind* but different in *direction*. Therefore, if beliefs are not ‘states that must fit to the world’ then the ‘direction of fit’ account is in trouble.

In what sense do beliefs and desires have different ‘directions of fit’? Platts puts this distinction like this:

Beliefs aim at the true... beliefs should be changed to fit with the world, not *vice versa*. Desires aim at realisation, and their realisation is the world fitting with them (1979 pp. 256-7)

However, since beliefs and desires cannot literally have aims, this talk must be metaphorical<sup>5</sup> (Zangwill 1998 p. 175). A further account is needed.

For now let us focus on beliefs (desires are discussed in sections 5.5-5.8). The account of the direction of fit of beliefs that Smith suggests is that “a belief that *p* tends to go out of existence in the presence of a perception with the content that not *p*” (1994 p. 115). However, beliefs cannot be *identified* by their tendency to go out of existence upon the perception that their content does not match the world, because “it isn’t clear that an individual belief is the sort of thing that could have such a tendency” (Shafer-Landau 2003 p. 35). We might be able to specify, given the particularities of the person in question, whether any given belief would persist in any given set of circumstances. But it is not clear that this describes a tendency. Presumably tendencies are dispositional properties, and in predicting when a particular agent would have or lack a particular belief, we are not picking out a dispositional property of the belief. Perhaps we are picking out a dispositional property of the agent, but if so, we are only picking out a disposition of the agent towards any belief with this content (i.e. this single belief) rather than of beliefs in general. Anyway, it simply isn’t true that, with respect to particular beliefs, or beliefs in general, agents are always disposed to relinquish beliefs upon perceptions that *prima facie* suggest their falsity. For one thing, people sometimes have irrational beliefs. As Shafer-Landau says, the fact that most beliefs extinguish upon a perception<sup>6</sup> whose content is inconsistent with the content of the belief “does not give us any basis for making a determination in particular cases as to whether a person believes that *p*, as opposed to desires that *p*” (2003 p. 35). He continues:

For beliefs sometimes do persist even after the agent recognizes that the belief is false<sup>7</sup>. And desires sometimes do extinguish after a perception that their content is unrealized... A man who perceives all of the excellent evidence for the straying affections of his wife may nevertheless continue to believe in her

---

<sup>5</sup> Geoffrey Scarre has suggested to me that this talk might be metonymic; it might be a way of saying that it is subjects who aim to fit to the world in believing and aim to change the world in desiring. However, there are problems with this suggestion that mirror the problems with seeing such talk as metaphorical. Some agents are irrational and irresponsible in their beliefs, and some are apathetic about pursuing their desires. We need to characterise beliefs and desires in normative terms (see below).

<sup>6</sup> It might seem that a belief that *p* would always extinguish upon perception with the content that *not-p* except in cases of extreme irrationality and delusion. However, imagine showing video footage filmed from a space shuttle to someone who believes the world is flat. They might continue to believe that *p*, the world is flat, despite this perception with the content that *not-p*, that the world is round. Note also that believing *p* whilst believing that *not-p* differs from believing the conjunction of these two claims (*p* and *not-p*). Believing a pair of contraries differs from believing a contradiction.

<sup>7</sup> This is not well phrased. Recognising that a belief is false seems necessarily to involve relinquishing that belief. I think Shafer-Landau means to point out that a belief can persist despite a perception whose content is inconsistent with the content of the belief.

fidelity... this man's attitude may have no tendency whatever to extinguish, despite his repeated perception that things are not as he takes them to be. Yet the attitude can be a belief for all that. (*ibid.*)

The problem with Smith's account is that it "fails to invoke normative standards" (*ibid.*). We need to talk of reasons or requirements rather than tendencies. Roughly speaking, we might suggest that for any belief that *p* there is a *pro tanto* reason to relinquish that belief upon a perception with the content that *q*, where *q* entails not-*p* or makes not-*p* significantly more probable. This reason may be outweighed. Plausibly, sometimes one should carry on believing that a good friend will not betray one's trust (*p*), despite a perception with content (*q*) that suggests<sup>8</sup> otherwise. There is still a *pro tanto* reason to relinquish *p*: one perceives that *q*, and *q* suggests that not-*p*. However norms of trust between friends can outweigh or defeat this reason.

These suggestions allow us to identify something shared by all beliefs: there are *pro tanto* reasons directing us to conform our beliefs to the world. Such requirements might be partially constitutive of beliefs and might help us to distinguish between beliefs and desires. However, this still gives us no reason to expect that beliefs and desires can be characterised as states with opposite directions of fit. We have seen no reason yet to expect the nature of 'desire' to ground *pro tanto* reasons to bring it about that *p* whenever one desires that *p*. More importantly, whilst the Humean argument is based on a non-normative conception of the 'direction of fit' of desires, we found we could only make sense of the 'direction of fit' of beliefs by thinking of it in normative terms. Thus, Humeans might end up with a notion of 'direction of fit' which means one thing (*pro tanto* reasons) in one instance (beliefs) and something else completely different (dispositions to action) in the other (desires). The attractive symmetry and simplicity of the 'direction of fit' account is completely lost; 'direction of fit' is not a unified concept, but an unstable chimera<sup>9</sup>.

## 5.5 The dispositional conception of desire

Although the notion of 'direction of fit' appears problematic, the argument for the Humean theory might still go through. Let us set aside our worries about the 'direction

---

<sup>8</sup> If the evidence is conclusive then one should *not*, but even if the evidence shows that it is more likely than not, arguably one should still trust that one's friend will nonetheless remain faithful.

<sup>9</sup> There is an easy way out of this: to adopt a normative conception of desire. However, I argue in chapter six that such a conception of desire cannot be used to establish the Humean theory.

of fit' picture of beliefs and, using Smith's characterisation of desires in terms of 'direction of fit', determine whether the Humean argument is valid.

Importantly, a substantial argument needs to be provided for the claim that desires can be characterised by their 'direction of fit' if it is to be used to establish the Humean theory. The point cannot be that we can adopt talk about states with which the world must fit as another way of discussing what we ordinarily call 'desires' and 'having a goal'. This would make the introduction of the 'direction of fit' terminology superfluous and misleading. The argument could go straight from 'desiring' to 'having a goal' (understood as 'having a motivating reason') without the intervening stage. However, the argument cannot proceed directly like this: this would represent an attempt to make the Humean theory true by definition. This would return us to Platts' dilemma, making the Humean theory *false* under the ordinary conception of desire and *trivial* under the new technical (pro attitude) meaning of desire. So the argument must introduce a new notion of desire that captures some interesting feature of our ordinary concept of desire but is defined independently of the notion of acting with a goal or for a motivating reason. The argument must then show how this new notion of desire features enlighteningly in explanations of action *qua* action – that is, in *rational* explanations of action that pick out what is distinctive about an event *as an action*. The form of explanation in question is that of giving a motivating reason. Schueler recognises this, saying:

[W]e have to understand 'world-to-word' direction of fit as referring to some feature of proper desires, which in the end, according to the argument Smith and Lenman propose, turns out to be what *accounts for* or explains goal directedness. (2003 pp. 28-9)

Humeans need to convince us that any state with a 'world-to-word' direction of fit is a desire and that no action can be explained without reference to a state with a 'world-to-word' direction of fit. What argument is offered for accepting this 'direction of fit' account? As Schueler points out, both Lenman (1996 pp. 291-2) and Smith<sup>10</sup> introduce the idea metaphorically:

To explain the world-to-word direction of fit by saying that 'the onus of match between the content of the thought involved and the world lies on the world and not the thought' is to give a misleading moral or legal metaphor (2003 p. 29)

<sup>10</sup> See Smith (1994 p. 112) where Smith picks up on Platts' (1979 pp. 256-7) discussion of the problematic nature of the metaphorical characterisation of 'direction of fit'.

Literally speaking there is no ‘onus’ on the world. The world has no *duty* to match itself to our desires. The same goes for examples. Saying that a wish to get to Cambridge leads me to change my direction of driving (Lenman 1996 p. 291, Schueler 2003 p. 29) (rather than the direction of my driving changing my wishes or beliefs) does not explain direction of fit. It purports to give an example of it. And, as Schueler points out, “I can... achieve a match between the direction in which I wish to travel and my actual direction *either* by changing my direction of travel *or* by changing where I wish to go” (2003 pp. 29-30). The claim that mental states can be categorised by their direction of fit is *presupposed* by such examples, not shown by them. Since the metaphor is unhelpful, a non-metaphorical account of ‘direction of fit’ is required.

In the case of desire, Smith attempts to cash out the ‘direction of fit’ metaphor using a dispositional conception of desire. Smith takes his lead from Hume’s suggestion that the calm passions “are more known by their effects than by their immediate sensation” (Smith 1994 p. 113). Remaining neutral on whether desires are causes, Smith endorses the claim that “desires are states that have a certain functional role” (*ibid.*). He claims that:

[W]e should think of desiring to  $\Phi$  as having a certain set of dispositions, the disposition to  $\Psi$  in conditions C, the disposition to  $\chi$  in conditions C', and so on, where, in order for conditions C and C' to obtain, the subject must have *inter alia*, certain other desires, and also certain means-ends beliefs, beliefs concerning  $\Phi$ -ing by  $\Psi$ -ing,  $\Phi$ -ing by  $\chi$ -ing and so on. (*ibid.*)

This allows that agents can be fallible about their desires, allows us to explain how desires have propositional content and gives us an epistemology of desire. And it “enables to cash out in non-metaphorical terms... the metaphorical characterization of beliefs and desires in terms of their different directions of fit” (Smith 1994 p. 115). This cashing out explains the difference between beliefs and desires as a “difference in the functional roles of belief and desire” (*ibid.*). Upon it being perceived that not-p, a belief that p tends to go out of existence whereas a desire that p tends to persist, disposing the holder to bring it about that p. Importantly, we already have a good reason to reject this account because we have already seen that this way of characterising beliefs is unsustainable.

According to Miller (2003 pp. 277-9) Smith’s argument for the dispositional conception of desire is “at best highly inconclusive” (2003 p. 270). Miller presents three

reasons for thinking so that mirror Smith's own objections to the phenomenological conception of desire: the dispositional conception of desire cannot make sense of the phenomenology, the epistemology or the propositional content of desire. I argue here that two of Miller's responses are inconclusive but that the third represents a good reason to reject the dispositional conception of desire. I start with the argument relating to the phenomenology of desire.

Smith claims that the fact that some desires have phenomenological content essentially, some have it contingently, and others lack it altogether, can be explained as follows:

According to [the dispositional conception], desires have phenomenological content just to the extent that the having of certain feelings is one of the things that they are dispositions to produce under certain conditions (Smith 1994 p. 114).

However, Miller (2003 pp. 277-8) objects, saying that a desire *produces* feelings seems equivalent to saying that it *causes* them to arise, and Smith states explicitly that his argument for the Humean theory is independent of the debate of whether reasons explanations are causal explanations (see Smith 1994 pp. 101-4). If the manifestations of desires are caused by those desires then actions are caused by desires in conjunction with means-end beliefs: reasons are causes. Therefore, since Smith ignores the debate about whether reasons are causes, he cannot presume to have established the Humean theory by the arguments he gives. However, given that we are willing to accept that reasons explanations are causal explanations in a minimal sense (see section 4.4), this conclusion gives us no reason to reject the dispositional conception of desire.

Concerning the epistemology of desire, Miller writes:

Consider my claim that I desire to finish this book by the weekend. I do not *infer* the fact that I have this desire from some hypothesis about how I will behave under contrary-to-fact conditions: I know it immediately and non-inferentially (2003 p. 278)<sup>11</sup>.

As Miller says, someone's claim to have a particular desire can be overturned by showing that they are victim to self-deception, but their claim is the default position. As such the epistemology of desire is non-inferential and first person authoritative. However, Miller thinks, the dispositional conception of desire entails that all knowledge



of desires, including one's own desires, is through inference from patterns of behaviour. Therefore the dispositional conception of desire is false since it entails, falsely, that the epistemology of desire is inferential and third personal even in cases of attributing desires to oneself. Miller's argument relies on the questionable assumption that one's dispositions cannot be known by introspection. Plausibly, exercise of the imagination can provide insight into what one would be disposed to do in counterfactual circumstances. Furthermore, even if desires do not *essentially* have phenomenological content, *feeling* disposed to  $\Phi$  might often be a reason for thinking that one desires to  $\Phi$ . Even if desires are essentially dispositions, self-knowledge and other contingent features of desires might facilitate non-inferential first person authoritative knowledge of one's desires. Therefore, the dispositional conception of desire might at least partially evade Miller's objection, and the objection seems inconclusive.

Miller's third objection is that it is not straightforward for the dispositional conception of desire to account for the propositional content of desires, because there are "serious... unresolved problems faced by dispositional theories of content determination" (2003 p. 279 cf. Kripke 1982). There are an infinite number of truths about cases in which it would be appropriate to use the term 'horse'. However, "the totality of my dispositions is finite, being the dispositions of a finite being that exists for a finite time" (Boghossian 1989 p. 509). Therefore, according to Kripke (1982), facts about how I am disposed to use a term cannot capture what I mean by that term. What is more, "if I mean something by an expression, then the potential infinity of truths that are generated as a result are *normative* truths: they are truths about how I *ought* to apply the expression" (Boghossian 1989 p. 509). Therefore, how I am *actually* disposed to use a term may mislead us about its meaning: I may not apply the term correctly.

The same considerations apply to desires. Smith suggests that the content of a particular desire is determined by what that desire leads a person to do, its manifestations. But, given the diversity of circumstances that we might possibly find ourselves in, and the huge variety of different means to achieving the same end, there is no end to the actions that would be appropriate means to the satisfaction of a given desire. For example, there is no end to the possible actions that may be ways to quench my thirst in different circumstances<sup>12</sup>. Since "the totality of my dispositions is finite,

---

<sup>11</sup> Davidson made a similar point forty years earlier (1963 p. 698).

<sup>12</sup> If this point is difficult to see, imagine different contraptions that could be set up to deliver water to a subject when a series of steps are followed, such as pulling levers, pushing buttons and turning handles.

being the dispositions of a finite being that exists for a finite time” (Boghossian 1989 p. 509), my desire to quench my thirst cannot be composed of dispositions to perform actions that are means to quench my thirst. The number of such possible actions is infinite. We might doubt that it really is impossible for us to have an infinite number of dispositions. After all, it might be appropriate to ascribe to people an infinite number of beliefs: ‘I believe that the number 3 is smaller than the number 4,’ ‘... than the number 5,’ ‘...than the number 6’ etc. *ad infinitum*<sup>13</sup>. Nonetheless, since we cannot have knowledge of an infinite number of dispositions, when I say ‘A desires to  $\Phi$ ’ I cannot *mean* he is disposed to  $\Psi$  in C, to  $\chi$  in C’ etc. *ad infinitum* – otherwise no one could ever be justified in their claims about anyone’s desires (even their own). And Kripke’s second point still applies: we cannot *identify* desires by what they dispose us to do because we may fail to pursue what we desire.

Hence, it is understandable that, in response to Smith’s claim that “the propositional content of a desire may then simply be determined by its functional role” (1994 p. 114), Miller remarks:

[T]here is nothing ‘simple’ about it. Smith is not entitled to the claim that the dispositional conception of desires can account for their propositional content. (2003 p. 278)

The whole appeal of explaining actions in terms of desires was that desires were thought of as having propositional contents that could relate straightforwardly to the outcomes of actions. However, if the Humean theory is dependent upon the dispositional conception of desire, and the latter cannot account for the supposed propositional content of desires, this appeal is completely lost. Moreover, any appeal that the dispositional conception of desire itself may have had is also lost. There then seems no reason to accept the Humean theory, and very good reason to reject the dispositional conception of desire.

## 5.6 Desire and ‘direction of fit’

There are further reasons for rejecting the dispositional conception of desire. The first is a more general objection against any attempt to characterise mental states in terms of

---

Through reflection on the endless possibilities in such contrived situations, the endlessness of the possibilities that could arise in more natural situations also becomes clear.

<sup>13</sup> Thanks to Geoffrey Scarre for this point.

their causal connections or functions<sup>14</sup>. The Humean argument requires that we explain goal-directedness of desires, of those states with ‘world-to-word’ direction of fit, “in a way that does not itself make use of notions that already involve goal directedness” (Schueler 2003 p. 37). This is because the Humean claims we should accept their theory because it is the only way to make sense of the goal-directedness of action (Smith 1994 p. 125). Moreover, since the Humean claims to explain the goal-directedness of action in terms of the direction of fit of desires, they cannot succeed if their analysis has to appeal to goal-directedness. Then the conclusion would be that to have a goal is to have a goal. Hence, the Humean appeals to a functional theory of mental states. Smith writes:

[T]he difference between beliefs and desires in terms of direction of fit can be seen to amount to a difference in the functional roles of belief and desire. Very roughly, and simplifying somewhat, it amounts, *inter alia*, to a difference in the counterfactual dependence of a belief that *p* and a desire that *p* on a perception with the content that not *p*: a belief that *p* tends to go out of existence in the presence of a perception with the content that not *p*, whereas a desire that *p* tends to endure, disposing the subject in that state to bring it about that *p*. (1994 p. 115)

However, it seems very unlikely that a functional theory of mental states would have the character that Smith supposes. As Schueler says:

Any functional account of mental states, or of features of mental states, will have to take as data... a detailed story about the roles of these states *vis à vis* all the other states in question. That story will be the basis on which it is decided which dispositions get built into which mental states. But this means that the truth or falsity of the Humean Theory of Motivation will have to be decided *before* any such functional theory gets written... It cannot be used to support the Humean theory. (2003 p. 35)

Since some beliefs – such as moral beliefs – seem to dispose people to action, it would seem that a functional account of such beliefs would have to take this into account. In writing our functional theory of mental states, we will need to decide whether or not moral beliefs do indeed dispose people to action. The Humean who wants to base their

---

<sup>14</sup> In fact there are two other strategies for accounting for ‘direction of fit’. One is to base the ‘direction of fit’ claim in an analysis of the inherent features of mental states. This strategy does not look promising. For one thing, it might be that mental states can only be understood holistically (see e.g. Davidson 1970). Moreover, according to Schueler an inherent features account would entail, falsely, that there is “a lawful connection” between mental states and actions (2003 pp. 30-1 cf. Schueler 2003 ch. 1). There is a third, normative, reading of ‘direction of fit’ (Zangwill 1998, Shafer-Landau 2003 p. 136) mentioned above (section 5.4) which is far more plausible. This is discussed in chapter six where I claim it cannot be used to establish the Humean theory.

theory on a functional theory of mental states cannot appeal to the Humean theory as a reason for not taking the motivational efficacy of moral beliefs at face value. Since the functional theory has not yet been written we do not yet know whether it supports the Humean theory. Hence, since the Humean theory cannot legitimately be used to shape the functional theory, the functional theory fails to support the Humean theory.

## 5.7 Desires, dispositions and counterfactuals

Since the dispositional conception of desire is a functional theory, it is threatened by the objections of the previous section. However, there are also independent reasons for rejecting the dispositional conception of desire. The Humean wants us to accept their thesis on the basis of the three following analyses:

- (1) Desires are to be characterised by their ‘direction of fit’;
- (2) The metaphor of ‘direction of fit’ is cashed out by giving a dispositional conception of desire (i.e. desires are dispositions)
- (3) Dispositions are to be analysed in terms of counterfactuals

However, I now argue that we should not accept (3) that dispositions can be analysed in terms of counterfactuals. Nor should we accept the second claim – that desires are dispositions – when it is conjoined with the Humean belief-desire thesis. Moreover, this is for reasons completely independent of those given earlier in section 5.3.

After announcing that we can understand “desiring to  $\Phi$  as having a certain set of dispositions [such as] the disposition to  $\Psi$  in conditions  $C$ ” (1994 p. 113)<sup>15</sup> Smith declares that “the epistemology of desire is simply the epistemology of dispositional states – that is, the epistemology of such counterfactuals” (*ibid*). However, it is unlikely that dispositions can be adequately analysed in terms of counterfactuals<sup>16</sup>. To analyse dispositions in terms of such counterfactuals is to say that having a disposition entails the truth of a conditional, such as: “If this glass had been suitably struck, then this

---

<sup>15</sup> Earlier authors have made similar claims. Wright follows Taylor in maintaining that the connection between desiring to  $\Phi$  and  $\Phi$ -ing under appropriate conditions is one “logical necessity” (Wright 1976 p. 119); desiring something *is* being disposed to try get it (Taylor 1964 p. 38). Wright assumes that action explanations are “explanations in terms of purposes, intentions, desires, and the like”, both teleological and given in terms of dispositions (1976 p. 121). Davidson, however, recognises that such claims are problematic since there is more to desires than their rationalisation of actions “– their expression in feelings and in actions they do not rationalize, for example” (Davidson 1963 p. 698).

striking would have caused the glass to shatter’’ (Armstrong et al. 1996 p. 15). This is Armstrong’s proposed analysis of what it means for a glass to be brittle. This analysis is problematic. Martin (1994) invites us to consider a wire connected to an ‘electro-fink’, a machine that can reliably detect when the wire touches a conductor. We are asked to imagine that the system is arranged so that when a conductor touches the wire, the electro-fink instantaneously makes the wire live, but that otherwise the wire is dead (pp. 2-3). Since the wire is dead, it lacks the disposition to transfer electrical charge to a conductor. Nevertheless, whenever a conductor contacts the wire, current flows from the wire, due to the electro-fink. The situation can be reversed:

We turn a switch on our electro-fink so as to make it operate on a reverse cycle, as it were. So the wire is dead when and only when a conductor touches it. At all other times it is live. (Martin 1994 p. 3)

The wire is live, having a disposition to transfer electrical charge to a conductor. Nevertheless, when a conductor contacts the wire, no current flows from the wire. Hence, the fact that current flows from the wire when it is touched by a conductor is neither necessary nor sufficient for the wire being live. However, when the wire is live it is disposed to transfer charge to conductors. This being the case, the dispositional property in question cannot be adequately analysed in terms of counterfactuals (Martin 1994; cf. Bird 1998 p. 227). As Fara says, “accounts which state the truth conditions of disposition ascriptions in terms of conditionals... are mistaken” (2005 p. 45); “no conditional account can reflect the fact... that dispositions may fail to become manifest even when their conditions of manifestation obtain” (p. 46).

Lewis (1997) attempts to overcome this problem by analysing dispositions in terms of intrinsic properties that provide the causal basis for whatever the manifestations of that disposition are. He claims that in finkish cases the intrinsic property that provides the causal basis is lost, and hence the disposition does not persist during the time of the stimulus. However, even if Lewis’ reply works against the initial argument offered by Martin, it cannot account for the possibility of antidotes:

When the sorcerer protects his glass, his strategy is finkishly to remove its fragility as soon as it is struck. Another way of protecting the glass once it is struck is to find an antidote to striking. The sorcerer, being a brilliant physicist, may be able to administer shock waves to the struck glass which precisely

---

<sup>16</sup> Thanks to Alex Carruth for suggesting this line of argument and pointing me in the direction of the work of C. B. Martin.

cancel out the shock of the original striking, hence saving the glass from destruction... the point of the antidote case, unlike the finkish case, is that the disposition remains. (Bird 1998 p. 228)

Similarly, Lewis' analysis cannot account for cases of 'masking' in which the disposition is not removed, but its manifestation is not produced because the intrinsic property that produces it is blocked. Johnston gives the example of wrapping a glass in bubble-wrap to prevent it breaking when dropped (1992 p. 233). The glass remains fragile, but its fragility is masked. As Fara points out, fragility is not lost because of extrinsic protection: "After all, the support structure is there precisely because the cup is disposed to break when struck." (2005 p. 49)

It is tempting to try to rescue the analysis of dispositions by adding a *ceteris paribus* clause. We might say that the fragility of a glass can be understood in terms of the fact that it would break when dropped, *other things being equal*. This, however, would make the analysis vacuous:

[T]o say that the glass vase would, *ceteris paribus*, shatter if it were dropped seems to be simply to say that if it were dropped then, unless it didn't shatter, it would shatter. (Fara 2009; cf. Martin 1994; Bird 1998; Mumford 1998; Molnar 2003 ch.4; Fara 2005)

Furthermore, there is no way of accounting for all the possible conditions that might mask a disposition. Such conditions are highly varied; "the only relevant property that cases of masking have in common with each other is that they are cases of masking" (Fara 2005 p. 51). To specify that dispositions manifest except when there are masking conditions would also be trivial. We were looking for an account of the manifestations of dispositions in terms of conditionals. To say that they manifest except where something stops them manifesting is not a way of giving such an account.

If dispositions cannot be analysed in terms of conditionals then, supposing that desires are dispositions, desires cannot be analysed in terms of conditionals. Therefore, Smith's proposed analysis is wrong. Smith claims that a desire to  $\Phi$  is a set of dispositions:

the disposition to  $\Psi$  in conditions  $C$ , the disposition to  $\chi$  in conditions  $C'$ , and so on, where, in order for conditions  $C$  and  $C'$  to obtain, the subject must have *inter alia*, certain other desires, and also certain means-ends beliefs, beliefs concerning  $\Phi$ -ing by  $\Psi$ -ing,  $\Phi$ -ing by  $\chi$ -ing and so on. (Smith 1994 p. 113)

The epistemology of the desire to  $\Phi$  is supposedly the epistemology of the counterfactuals ‘A would  $\Psi$  in conditions C’, ‘A would  $\chi$  in conditions C’, and so on. There are two claims here: (i.) that the desire to  $\Phi$  should be analysed in terms of the dispositions in question (ii.) that these dispositions should be analysed in terms of the counterfactuals, or conditionals, in question. Putting the parts together gives an analysis of desires in terms of counterfactuals. Whether it is desires or dispositions that are being analysed in terms of conditionals, the same point applies. The analysis cannot be successful because there are things that can interfere, either systematically or occasionally, with the disposition’s tendency to bring about its manifestations.

In the case of desires or dispositions of an agent, the agent themselves can play the role of the ‘electro-fink’ or the ‘mask’. For example, an agent may be aware of having an immoral desire. Consider a man who is aware that he finds the idea of cheating on his wife with a particular woman intrinsically attractive – being unfaithful is part of what he desires, he finds this intrinsically attractive. Suppose, however, that he resists temptation, and that nothing could induce him to give in. The man’s desire to be unfaithful cannot be analysed in terms of what he would do given certain conditions, for he deliberately regulates his conduct so as to prevent the desire from having the manifestations in action that it disposes him towards. Suppose we accept that this desire is a disposition to unfaithfulness. We might say that this desire is the kind of desire that usually disposes people to unfaithfulness or that it would dispose him to unfaithfulness if he were not so good at resisting temptation. We might still wonder whether the desire *actually* disposes him to unfaithfulness and whether or not he, overall, is disposed to be unfaithful.

In cases of agency, where choice plays an essential role, talk of dispositions might seem inappropriate since it suggests fixed properties that determine action. This shows that dispositions cannot play the explanatory role that the Humean requires: a disposition only explains an action on the assumption that an agent does not mask, finkishly remove or provide an antidote, to the disposition. As well as appearing vacuous, this analysis amounts to saying that *actions* are explained only in cases where the *agent* is *passive*, and this is absurd. Surely action explanations should work best when, and tend to bring out the fact that, an agent directly and actively brings it about, via the workings of choice and rational agency, that they  $\Phi$ . Moreover, there remains the problem that desires (e.g. the desire to unfaithfulness) cannot be analysed in terms of counterfactuals as Smith supposes. Although the man’s desire may manifest itself in

his behaviour – he may avoid places where he will meet the woman to whom he is attracted, for example – the desire will not have the right kind of manifestations. Smith wants us to analyse a desire to  $\Phi$  in terms of dispositions to do things that the agent believes are ways to  $\Phi$ . The man's desire to be unfaithful has manifestations, but it has no manifestations of *this* kind. He deliberately prevents it from having any such manifestations, acting like an 'electro-fink'.

The analysis of desires in terms of dispositions-as-counterfactuals does not succeed. This failure is instructive. One of the reasons it fails is because agents are, to some extent, free to decide what to do about their mental states: whether to retain them, use them in reasoning, act on them, and so on<sup>17</sup>. Any account that seeks to characterise mental states and behaviour in terms only of the functions and interrelations of those states misses something vital: the role of the agent herself.

## 5.8 Passive action

Humeans cannot characterise desires in terms of 'direction of fit'. However, I now argue that even if the functional analysis had succeeded, we would not have ended up with a satisfactory account of action since the Humean-dispositional analysis reduces action to something passive, to a form of mere activity in which the agent is carried along. It does not allow us to preserve what is distinctive about action by making sense of the agent *qua* agent as active, as one who controls, decides, and instigates.

According to Humeans, agents are morally good if they have moral desires, moral dispositions. However, this seems to mean that moral goodness or virtue is something outside the agent's active control, something with respect to which the agent is *passive*. If having certain dispositions constitutes moral perfection, one might be a morally perfect agent by accident, just by happening to be fortunate enough to have certain dispositional properties. Just as a wineglass is disposed to break when struck, so morally perfect people are disposed to act rightly when an opportunity for moral action arises. Given the relevant stimulus, the response is elicited automatically by the entity with the dispositional property. On the Humean-dispositional account, agents who have achieved moral perfection are like this: the right stimulus suffices to elicit the relevant

---

<sup>17</sup> To suggest that the agent must have other mental states that enable her to decide what to do about these mental states is either to invite an infinite regress or to return us to the problems just discussed. However, I do not see why we should have to make such a suggestion.



behaviour. Agents who are morally mediocre are also like this; it's just that the propositional contents of their dispositional states differ.

This seems like an inadequate account of moral character. Moral perfection is something to be praised and admired and something that people have to constantly strive towards. Being morally good is often a struggle; people are responsible for their moral character. Moral perfection is not a state that one could find oneself in by accident, and nor is moral depravity. Moreover, our moral character does not *determine* our actions. Although our moral character can be used to explain or predict our actions, we are free to act in ways that are out of character and, in so doing, to reshape our character. If this were not so, after committing evil it would be a valid excuse to point out one's moral depravity – that all of one's dispositions happened to be towards committing acts that are evil. The Humean-dispositional account is supposed to capture the essential nature of action, yet apparently it fails to account for the fact that agents are, to a significant degree, responsible for their own actions and character. Furthermore, it characterises the difference between moral perfection and moral depravity as a difference between the propositional contents of the dispositional properties of agents. This is not an adequate account of moral agency or of the difference between vice and virtue, depravity and perfection. It does not allow us to say anything about the extent to which people are morally responsible for their character<sup>18</sup>.

Smith might reply as follows. He might point out that, as a matter of contingent fact, no one is born with the dispositions constitutive of moral perfection. Becoming morally good requires undergoing moral development and education and shaping our dispositions through choice, discipline, reflection, exercise of the imagination, and so on. Therefore, whilst people might be passive with respect to, and not responsible for, some of their dispositions, they are responsible for many of them. Therefore, Smith may say, his account does allow that people can be held responsible for the extent to which they approximate to the ideal of moral perfection.

However, the Humean-dispositional account undermines this response. To the extent that one is responsible for shaping one's own character or dispositions, one must shape those dispositions intentionally. According to the Humean, anything done intentionally is done for a motivating reason consisting of a belief-desire pair. Therefore, if desires are dispositions, one only shapes one's dispositions in ways that

---

<sup>18</sup> It also seems phenomenologically inadequate, but I focus here on the notion of responsibility.

one is disposed to shape them. Every shaping of a disposition must be explained in terms of a previous disposition. Ultimately this explanation will terminate with dispositions that the agent had at birth. Therefore, agents cannot be held responsible for their dispositions, for their moral character, since they cannot be held responsible for the dispositions they were born with. Thus, the dispositional conception of desire, taken together with the Humean theory, entails that we cannot be responsible for any of our actions<sup>19</sup>. But since we are responsible for many of our actions, the Humean theory, when interpreted according to the dispositional conception of desire, is false.

This problem is an instance of a problem that Velleman (2000) identifies with the following idea, which he calls ‘the standard story of human action’:

[W]hen someone acts... [t]here is something that the agent wants, and there is an action that he believes conducive to its attainment. His desire for the end, and his belief in the action as a means, justify the action, and they jointly cause an intention to take it, which in turn causes the corresponding movements of the agent’s body (2000 p. 123).

Velleman’s complaint is that the standard story of human action “fails to include an agent – or, more precisely, fails to cast the agent in his proper role” (2000 p. 123). This story is one in which “events take place inside a person, but the person merely serves as the arena for these events: he takes no active part” (*ibid*). This is illustrated strikingly by the fact that “an agent’s desires and beliefs can cause a corresponding intention despite him, and hence without his participation” (Velleman 2000 p. 125, cf. Frankfurt 1988 pp. 18-21). Frankfurt describes a drug addict who is moved to take a drug by a force external to his own agency, yet whose action is describable (if anything is) in terms of his beliefs and desires (dispositions). This shows that the Humean-dispositional picture fails to distinguish between actions that show the agent in her role as rational animal, demonstrating the workings of agency, and those that demonstrate a failure of agency, when the agent acts despite herself.

Humeans cannot recapture this distinction by pointing to the difference between the presence and absence of a good reason for action. Compare the actions of a drug addict with those of a charity worker who tries to help the addict to overcome his addiction. According to the Humean, both should be understood in terms of belief-desire pairs. However, the former case demonstrates a failure of agency whereas the

---

<sup>19</sup> The only way to escape from this is by introducing normativity into the picture e.g. by accepting a normative conception of desire. This is considered in chapter six.

latter instantiates the proper workings of agency. The Humean explanation does not capture this important distinction. The Humean can point out that the charity worker believes that what they are doing is valuable whereas the drug addict does not. But this doesn't capture the difference in right way. The relevant difference is not in something that merely *accompanies* agency; it is a difference in agency itself. What the account needs to capture is the difference between whether one actively and responsibly determines one's behaviour or whether one is swept along by events inside one that one did not initiate and does not even endorse. The mere presence or absence of an accompanying evaluative belief cannot explain *that* kind of difference. This is why Velleman says that the desire-belief model "is a model of activity but not of action" (2000 p. 10).

A general problem with the Humean model is that it doesn't allow us an adequate account of the way that agents take initiative when they act. Although the Humean account does not characterise agents as completely passive with respect to their mental states, it characterises them as passive with respect to them when it comes to initiating actions. The Humean encourages us to think of intentions as produced by agents from mental states that exist prior to the deliberative process. There is nothing – or almost nothing<sup>20</sup> – which the agent herself does when she acts or decides to act. Her role is restricted to putting together two pre-existing items whose contents predetermine that they fit together<sup>21</sup>.

Wallace proposes that we need "to expand our conception of the basic elements involved in reflective agency" (2006a p. 176) to include volition, which includes states that are "directly under the control of the agent... intentions, choices, and decisions" (*ibid.*). Wallace writes:

[I]ntentions, decisions, and choices are things we do, primitive examples of the phenomenon of agency itself. It is one thing to find that one wants some chocolate cake very much, or that its odor reminds one of one's childhood in Detroit, quite another to resolve to eat a piece. The difference, I would suggest, marks a line of fundamental importance, the line between the passive and the active in our psychological lives (2006a pp. 176-7).

---

<sup>20</sup> Smith does say that "in order for a desire and belief to constitute a motivating reason the agent must, as it were, put the relevant desire and belief together" (1994 p. 92). I discuss this idea in chapter six.

<sup>21</sup> It may seem that I am being unfair to Humeans by criticising their position without offering an alternative. However, I develop my own account of agency in chapter eight.

The Humean theory ignores this distinction, reducing the active to the passive and leaving us unable to understand how agents are responsible for the moral character of their actions or characters: in reducing the active to the passive, it reduces agency to mere behaviour or mere activity<sup>22</sup>. Thus, Velleman writes, “the standard model is sufficient for motivated activity but not for autonomous action” (2000 p. 10). When Anscombe says: “The primitive sign of wanting is *trying to get*” (1957 § 36) it is instructive that she immediately goes on to say:

[I]n saying this, we describe the movement of an animal in terms that reach beyond what the animal is now doing. When a dog smells a piece of meat that lies the other side of the door, his trying to get it will be his scratching violently round the edges of the door and snuffling along the bottom of it and so on. Thus there are two features present in wanting; movement towards a thing and knowledge (or at least opinion) that the thing is there. When we consider human action, though it is a great deal more complicated (*ibid.*).

The reason why the human case is more complicated is related to the fact that “the human mind is self-conscious in the sense that it is essentially reflective” (Korsgaard 1996b p. 92). We can reflect on our desires; even when we have “a powerful impulse to act” we can “bring that impulse into view” and question whether it is “really a reason to act” (Korsgaard 1996b p. 93). It is this that gives rise to what Korsgaard calls “the most striking fact about human life” (1996b p. 1): the fact that we, unlike a dog snuffling around a door, have values. The Humean theory does not allow us to make sense of this ‘striking fact’ since it does not allow us to see human beings as autonomous, as actively deciding what to do, and as capable of acting for good reasons.

## 5.9 Conclusions

In this chapter we have examined Smith’s argument for the Humean theory. We have seen that the first premise of Smith’s argument is true so long as it is not interpreted according to the teleological concept of action. Actions do not necessarily aim at some ‘end state’. Secondly, it was argued that performing an action doesn’t necessarily involve having a belief about how to perform that action. Knowing-how to  $\Phi$  is not reducible to knowing-that  $w$  is a way to  $\Phi$ . Hence, the belief-desire thesis is false. The notion of ‘direction of fit’, introduced in the second and third premises of Smith’s

---

<sup>22</sup> Humeans might respond by claiming that there is no reason for thinking that we are passive with respect to our dispositions. I claim in chapter six that this response leads directly to an anti-Humean

argument, was examined. It was claimed that in the case of beliefs ‘direction of fit’ could only be a normative relation. This undermines Smith’s argument since it shows that ‘direction of fit’ cannot be a unified notion playing the role that Smith requires. The ‘direction of fit’ of desires was then discussed. It was concluded that functionalism, an inherent features account and the dispositional conception of desires all failed to provide a plausible account of direction of fit. Moreover, it was claimed that the dispositional conception of desire advanced by Smith is independently implausible and leads to implausible results when combined with the Humean theory. Overall, it is concluded that Smith’s argument for the Humean theory fails.

## Chapter 6. Normativity and the Humean Theory

### 6.0 Introduction

Chapter five contained a lengthy critique of the Humean claim that desires can be characterised by their ‘direction of fit’ to the world. I claimed that the belief-desire thesis is false, that action does not necessarily aim at an end state, that the dispositional conception of desire is false, and that the Humean theory would be seriously inadequate even if it could be established in the way Smith imagines. It might be wondered whether this was worthwhile, given that a more plausible conception of ‘direction of fit’ is available: a normative conception (Zangwill 1998). According to Shafer-Landau, Smith can alter his argument by using the normative conception of ‘direction of fit’ “at no cost to the master argument that he is advancing” (2003 p. 136). This would overcome some (although not all) of my previous objections to Smith’s argument. Therefore I need to examine this normative conception of ‘direction of fit’ and consider what the argument for the Humean theory will look like if it is adopted.

### 6.1 The normative reading of ‘direction of fit’

Shafer-Landau explains this conception of ‘direction of fit’ in the following way:

Those who perceive that their goals aren’t yet satisfied, and then sit by and do nothing, don’t really have goals at all; just wishes, or fantasies, or hopes... we do best to replace Smith’s talk of tendencies to endure with explicitly normative criteria... beliefs, for instance, are states that *ought to* extinguish, rather than those that tend to extinguish, upon a (non-discounted) perception that their content is not realized... the appropriateness of a goal isn’t determined by whether it accurately represents the world as it is... In simple terms: the fact that the world doesn’t match the goal’s content doesn’t impugn the goal. (Shafer-Landau 2003 p.136)

This characterisation gets us started, but it simply doesn’t seem right that beliefs ought always to extinguish upon non-discounted perceptions that their content is not realised. We should not be quite so epistemically fickle<sup>1</sup>. And it is unclear whether Shafer-

---

<sup>1</sup> It is not fickle to change one’s mind when confronted with superior evidence, but being confronted with a single non-discounted perception that seems *prima facie* to show that not-p is compatible with there being better evidence that p and does not mean that we should cease to believe p. If we have good reasons to believe p in the first place then we ought to try to see whether or not there is any way that the

Landau is arguing that actively seeking to realise a goal is *constitutive* of having a goal or whether he is making the incompatible claim that there is a *rational requirement* for us to seek to realise our goals. Zangwill expresses the normative conception of desire more clearly:

The belief that p would be *pro tanto* rational if it were caused to go out of existence in the presence of the perceptual experience as of not-p. The desire that p would be *pro tanto* rational if it endured given the perceptual experience as of not-p; and if the desire that p caused the attempt to bring about that p then the attempt would be *pro tanto* rational. It could never be enough to say that beliefs and desires *have* certain causal roles with respect to other mental states, for those roles are rational roles; they are the roles they *ought* to have. (1998 p. 196)

If the normative conception of ‘direction of fit’ is to be plausible, it must be construed in terms of *pro tanto* (rather than overall) rational requirements in order to do justice to the complexity of normative requirements (cf. Dancy 2004 pp. 38-49). Seeing ‘direction of fit’ in this way involves seeing the presence of desires and beliefs as grounding contributory ‘ought’ claims<sup>2</sup>. On this account desires provide contributory *normative* reasons for action when taken together with justified<sup>3</sup> beliefs<sup>4</sup>. Supposing we accept this picture, what would be the implications for the Humean theory?

Humeans claim that rational requirements only explain actions when agents *desire* to conform to those requirements. This claim is immediately cast into doubt by characterising beliefs and desires normatively. If the relation between desire and action is that desires ground *pro tanto* rational requirements to act then acting on a desire, d, involves conforming to a rational requirement, r. According to Humeans, conforming to r must be explained by a belief that one is conforming to r and a desire, d<sub>2</sub> to conform to r. This desire, d<sub>2</sub>, in turn, relates to action by grounding a *pro tanto* requirement to act, r<sub>2</sub>. According to Humeans, conforming to r<sub>2</sub> must be explained by a belief that one is conforming to r<sub>2</sub> and a desire, d<sub>3</sub> to conform to r<sub>2</sub>. An infinite regress threatens. There is

---

perception that suggests not-p might be accommodated within a view that allows us to carry on believing that p. If we have already perceived several things that seem to show that not-p or we do not have very good reasons to believe that p then perhaps our belief *should* extinguish upon this new non-discounted perception that not p. The point is that such perceptions ground only *pro tanto* requirements (as argued, following Zangwill, in the main text).

<sup>2</sup> See Dancy (2004 pp. 15-37) for an excellent discussion of such contributory ‘oughts’ or contributory reasons.

<sup>3</sup> Or, perhaps, *true* and justified

<sup>4</sup> Beliefs, on the other hand, would be seen as grounding contributory theoretical normative reasons (i.e. reasons for having other beliefs).

another problem for Humeans as well: if desires ground *pro tanto* requirements to act then whether desires are motivationally efficacious is under rational control. If so, reason is the master of the passions, not its slave. Insofar as we are rational, *we*, as agents, through employing our rational faculties, can decide on the basis of available reasons whether to act on a particular desire or not. If Humeans admit this, there is little left of their theory. But it is difficult to see how Humeans can avoid these conclusions if they embrace a normative conception of desire.

The normative conception of desire leaves Humeans no room to claim that there cannot be rational norms that require we regulate our desires in particular ways. When we characterised mental states functionally and had no normative constraints in the motivational picture, this question could more easily be brushed aside. Now Humeans admit there is no general reason for thinking that rational norms cannot apply directly to motivational states since they recognise at least one such norm: that grounding a *pro tanto* requirement to realise the propositional contents of one's desires. Humeans now owe us a good argument showing why there cannot be other norms applying directly to motivational states. There certainly *seem* to be other such norms: we should seek to rid ourselves of immoral desires, for instance. A mere reiteration of the claim that desires ground *pro tanto* rational requirements with a particular direction of fit will not do. Even if this is supposed to be constitutive of desire<sup>5</sup>, it does not explain why other norms cannot apply to desires.

Similarly, the normative interpretation of 'direction of fit' has implications for the belief-desire thesis. We are no longer thinking of desires as dispositional states that manifest themselves in action whenever the right circumstances obtain but as states that ground *pro tanto* rational requirements to attempt appropriate actions. Circumstances, *C*, that, given the dispositional conception, we would expect to elicit a particular action,  $\Phi$ , on the basis of the presence of a desire, *d*, might ground no such expectation under the normative conception. The normative conception allows that agents may choose not to act upon desires they have. Flouting the requirement grounded by the desire is just one way they might do so. Supposing that they are subject to a number of other requirements as well, they might judge that overall *d* provides insufficient reason to  $\Phi$ , being outweighed by other considerations. On the dispositional conception of desire we also need to take into account competing dispositions, but the conditional analysis

---

<sup>5</sup> This claim has not yet actually been *substantiated*, and, if the arguments of chapter seven are right, it is false.



would always have allowed us to predict which dispositions would lead to action under which conditions. On the normative conception, rational acts of judgement play an irreducible role.

However, if we suppose intentions are formed through such rational acts of judgement we can ask why desires are necessary as inputs in this process. After all, the desire only plays a role through grounding a rational requirement. If rational requirements can be grounded in other ways then there is no reason why instances of intention forming judgement shouldn't occur without desires supplying the *pro tanto* requirements taken as input. We can also ask, if agents can choose not to act when desires are present why shouldn't they choose to act in the absence of a desire? The most obvious candidate case for such action without desire is when an agent mistakenly thinks they have a desire and hence (if they accept the normative conception of desire, at least) takes themselves to have a reason for action. Admitting this possibility shows us that the presence of desires is unimportant: what is important is that the agent chooses to act in a particular way because they take it to be required or favoured in some way. If we accept the normative conception of desire, thinking that an action would satisfy some desire that one takes oneself to have is one way – but only one way among others – that one can take one's action to be to some extent favoured in some way<sup>6</sup>. Therefore, the belief-desire thesis cannot be maintained, since actions can be motivated and explained without reference to any desire of the agent. They can be motivated and explained by the agent's conformity to rational requirements or by the agent taking the action to be favoured in some way. Therefore the normative conception of desire cannot be used to establish the Humean theory since once the Humean theory is interpreted according to the normative conception of desire it collapses into anti-Humeanism.

## 6.2 The assumption of means-end rationality

Until now we have been ignoring an important qualification to Smith's position. Smith claims that belief-desire descriptions only explain actions on the assumption that the

---

<sup>6</sup> It does not follow that all practical requirements are grounded in desires. In chapter seven I argue that if there are any rational requirements, none of them are provided by desires: the presence of a desire is never sufficient to ground a rational requirement.

agent in question is rational<sup>7</sup>, since otherwise they might fail to put the belief and desire together in the right way<sup>8</sup>. This makes the connection between the *presence* of belief and a desire with the right contents and the *event* of their actually motivating an action, a normative one for the following reason. Being rational consists of nothing other than conforming to rational requirements; if some proposition about something an agent did is true of an agent in virtue of their being rational it must be true that there was a rational requirement for them to do that thing. Therefore, if it is true that being rational entails that an agent will put together appropriate beliefs and desires into pairs that can explain actions, it follows that there is a rational requirement for agents to put their beliefs and desires together in this way. Rational requirements are normative<sup>9</sup> and therefore provide normative reasons for action. Therefore, Smith's claim entails that whenever agents have an appropriate related belief and desire they have a normative reason to put these together. This entails that, an agent who has a belief and desire with the appropriate contents *ought* to put them together: she *ought* to have a motivating reason and be in a state that is potentially explanatory of action. Smith's claim therefore seems to entail desiring to  $\Psi$  and believing that by  $\Phi$ -ing one can  $\Psi$  grounds a *pro tanto* normative reason that one  $\Phi$  and this entails that if there is no reason to do anything incompatible with  $\Phi$ -ing, one ought to  $\Phi$ . Since to ground such a contributory ought claim is to provide a normative reason, this now sounds like a theory of normative reasons or of rational action<sup>10</sup>. This is problematic, since Smith already has a (quite different) theory of normative reasons. It looks as if he might end up with two (conflicting) theories of normative reasons, and no theory of motivation. It would seem that the obvious way to resolve this is by adopting a minimal anti-Humean theory of motivation according to which motivation explanations of actions just involve mentioning what the agent took to be good reasons.

---

<sup>7</sup> Smith (2009) suggests that Davidson does not recognise this condition whilst Hempel (1961), a Humean contemporary to Davidson, does. Smith, on the other hand explicitly says, "the agent must, as it were, put the relevant desire and belief together" (1994 p. 92).

<sup>8</sup> Dreier (1997) also makes this point, claiming that means-end rationality is necessary for an agent to have any reasons at all. There is an interesting explanatory question here: is it cases when the presence of a belief and desire with the appropriate contents fails to lead to action that require explanation, or those in which it succeeds, or both? At this point Smith offers us no more of an explanation of their coming apart or not than to say that in one case the agent is rational in a particular way, and in the other she is not. In the current context there seems to be no reason to pursue this question, however.

<sup>9</sup> See chapter 1 footnote 7.

<sup>10</sup> Morton says, "when philosophers try to describe the theory [of motive and action] it collapses into the theory of rational action" (2007 p. 217). Smith's version of the Humean theory is in danger of suffering from just such a collapse.

There is another interesting consequence of Smith's claim that conformity to a norm of rationality is necessary to get from the presence of a belief and a desire with the right contents to actually acting on the reason they potentially constitute. Dreier (1997) points out that conforming to the norm of means-end rationality cannot be dependent upon any desire to conform to that norm. Such a desire could only motivate one to conform to the means-end norm if one is *already* means-end rational. However, if one can conform to the means-end norm independently of any desire to do so, why should one not also be able to conform to other norms independently of one's prior desires? If an assumption of means-end rationality is required for explanation of actions, why shouldn't other norms be able to play an irreducible, non-desire based role in these explanations as well?

A point of clarification is required: ends and desires are not the same. To adopt something as an end is to will, intend or choose it, or at least to plan to pursue it or adopt it as a project. If one plans or chooses to do something but takes none of the steps necessary for doing it then either one has changed one's mind or is irrational, defeating one's own purposes<sup>11</sup>. However, one can desire something without choosing it<sup>12</sup>, and hence *prima facie* there does not necessarily appear to be any irrationality in failing to take the means to what one desires<sup>13</sup>. Smith and Dreier think otherwise, but they need to provide an argument for their claim. I think it would be more accurate for Dreier to speak of the means-desire rule rather than the means-end rule, but nonetheless I continue to use latter term.

What reason could Humeans give for thinking that the means-end rule is normative and that it is the only rule that it is possible to conform to independently of one's desires? Humeans might claim that the fact that the means-end rule requires a desire as part of its input is relevant here. They might claim that only norms that have motivational states as inputs can have motivational states as outputs. However, we not only need a reason for supposing that this is true, Humeans also need to convince us that the desires that the means-end rule requires as input really *are* intrinsically motivational states. If desires cannot lead to action without means-end beliefs and conformity to the means-end rule then it seems that they are at best contingently and extrinsically

---

<sup>11</sup> This does not mean that merely adopting something as an end gives one a reason to pursue it: it gives one a reason to *either* pursue it *or* abandon it as an end.

<sup>12</sup> I argue in chapter seven that our desires do not, by themselves, give us normative reasons for action.

<sup>13</sup> It may often be irrational to desire what one thinks there is no reason to pursue, but this is a different matter.

motivating. Humeans deny that there are other norms that can lead directly to action without going via the means-end norm. But why shouldn't there be other norms that take contingently motivating states as their inputs and have action as their output? And why shouldn't they do so without the need for the agent to have any independent desire to conform to the norm? If conformity to the means-end norm can exhibit these characteristics, then why shouldn't conformity to other norms also do so? Moral judgements seem to be just the kind of contingently motivating states that could provide input for norms that direct action. Why shouldn't there be a norm directing us to do what we judge we are morally required to do, and why shouldn't that norm motivate us independently of any desire to conform to it?

Dreier might reply that only the means-end norm is such that were it not possible to conform to it independently of one's desires then one could not conform to it at all. This is true, but irrelevant. It does not even establish that one can conform to the means-end norm<sup>14</sup>. So it certainly does not establish that *only* the means-end norm can be conformed to independently of one's desire to so conform. Importantly, *if* the means-end norm can be conformed to independently of a desire to conform to it, this shows that there is no general principle ruling out conforming to norms independently of a desire to conform to the norm. Then the question naturally arises as to why conformity to *any* norm should be dependent on a desire to conform to that norm. Then, of course, the question arises as to why we should accept the Humean theory of motivation at all. If norms can motivate us independently of our having a suitably related desire, why shouldn't *any* consideration be able to move us independently of having a related desire? Once again, introducing normative considerations into the Humean theory of motivation causes it to collapse into a minimal anti-Humeanism.

### 6.3 Motivating beliefs

We have already concluded that Smith's argument cannot be used to establish the Humean theory no matter how we characterise desires. However, even if the argument were sound and established the belief-desire thesis, this might be a hollow victory for

---

<sup>14</sup> The fact that people tend to pursue what they desire does not provide any support for Dreier's position. As we saw in chapter four in relation to Hume's views, normative nihilism is a possibility. The fact that people tend to pursue what they desire does not show that they are following a norm in doing so. The claim that rationality requires taking the means to satisfying one's desires is a substantive position that needs to be argued for: since it is a normative claim it is neither a conceptual platitude nor an empirical observation.

Humeans. The belief-desire thesis might not be the interesting claim it first seemed. To see why, consider Shafer-Landau's distinction between two senses in which beliefs might be sufficient for motivation:

The first, stronger version has it that evaluative beliefs are all by themselves enough to get one going—no desires needed. The weaker view claims that beliefs must be supplemented by desires, but that the relevant desires can themselves be generated by evaluative beliefs. Beliefs are, all by themselves, sufficient to yield motivated or derived desires, which, together with the evaluative beliefs, generate motivation (2003 p. 139).

We have already seen that Humeans have given us no reason to suppose that the stronger claim is false, and we have good reasons to reject Humeans' own positive claims. Suppose Humeans accept that evaluative beliefs are sufficient for motivation in the weaker sense, generating desires that can explain action. Even this is enough to allow that moral judgements might be intrinsically motivating<sup>15</sup>. However, there is nothing to stop us from moving from this weak claim to the stronger anti-Humean claim. Shafer-Landau writes:

The anti-Humean may even deny that the desires given in belief-desire pairs have any psychological reality independent of the factors that produce them. Saying that an agent desires to  $\Phi$  may just be a way of saying that some belief or norm motivates them to  $\Phi$  (2003 p. 139).

We have already seen that an agent does not need to have any belief about how to  $\Phi$  in order to  $\Phi$  (chapter 5, § 5.3). Why should we necessarily have to posit a desire to explain why an agent, taking  $\Phi$ -ing to be favoured or required in some way, decides to  $\Phi$ ? When we say that John  $\Phi$ -ed *because* it was his duty, the 'because' here captures the fact that his awareness of his duty motivated him. There is no reason for thinking that desires must be present in such cases as psychologically real mental states. Moreover, no mention of a desire is necessary for explanation; adding a desire to the explanation does not increase its explanatory power at all since there is nothing for the presence of the desire to explain.

---

<sup>15</sup> Actually, in some places Smith advances something very much like the weaker claim: in Smith (2009) but also earlier in Smith (1994 chapter 5) when he discusses normative reasons. (I do not think we should accept Smith's theory of normative reasons either, anti-Humean though it is. I discuss normative reasons in chapters seven and eight.) When he discusses the Humean theory of motivation (1987; 1988a; 1994 chapter 4; 2003) it sounds as if he rejects both the weaker and stronger claims.

Shafer-Landau claims that in some cases we form intentions on the basis of taking ourselves to have a particular desire or goal, and knowing how to satisfy desires or achieve goals or that kind (2003 p. 139). This might be characterised as being motivated by a combination of two beliefs (Dancy 1993 ch.2; Dancy 2000a)<sup>16</sup>. This mode of explanation does not rely upon the agent being mistaken about their desires, and it does not depend upon one's beliefs being about one's own desires. As Shafer-Landau says:

Any case in which one reflects and deliberates about what to do will be best explained by one's (true or false) beliefs—beliefs about what is desirable, about what one desires, about what is worthy, important, right, valuable, good, or interesting to pursue, plus beliefs about how to get there. (2003 p. 140)

Shafer-Landau warns that we should not think of all action like this, since not all action is preceded by reflection and deliberation. We should also be wary of falling back into the teleological conception of action, or of thinking that one cannot act without having a belief about what one is bringing about. The important point is that the fact that one took things to be a certain way can function perfectly adequately as an explanation of why one acted. Whether or not this is so is sometimes put in terms of the question of whether there are 'besires', whether 'besiring' is possible (see e.g. McDowell 1978; Altham 1986; McNaughton 1988; Smith 1994, pp. 118–125; Zangwill 2008a). According to Smith, if there were any 'besires', they would be mental states that have both the characteristics that Smith attributes to desires, and the characteristics he attributes to beliefs. In Smith's terminology, this would amount to them having both 'directions of fit'. Thus Smith has a straightforward objection to the possibility of besires<sup>17</sup>:

A state with both directions of fit would therefore have to be such that, both, in the presence of such a perception [that not-p] it tends to go out of existence, and,

---

<sup>16</sup> Dancy explains this view, which he calls 'pure cognitivism', as follows: "a desire is never a necessary part of what motivates... we have two beliefs which together motivate. One of these is a belief about how things are, and the other is about how things would be if the action were successfully performed" (2000a, pp. 13-14). Dancy once accepted this view (1993) as the only alternative to a Humean account, but later came to reject all psychologistic accounts of motivating reasons (2000a; 2003a; 2003b).

<sup>17</sup> Bittner is somewhat uncharitable in declaring that Smith "needs to put forward additional considerations" (2001 p. 10 § 15) against the possibility of besires because "contrary to his suggestion, he does not reject it on the strength of the argument outlined in the last paragraph" (*ibid.*). That is to say, Bittner says that the argument Smith uses to establish the Humean theory of motivation does not establish that there are no besires. This is true, but Smith does offer additional arguments. The trouble is that none of these arguments seem to work.

in the presence of such a perception, it tends to endure, leading the subject who has it to bring it about that p. (1994 p. 118)

However, given that this conception of ‘direction of fit’ cannot be sustained, neither can this response. Furthermore, as Zangwill points out, even if this conception of ‘direction of fit’ could be sustained, the argument would not work:

Smith’s claims about beliefs and desires beg the question against besire theorists, for they will just deny that the perception that p tends to cause the besire that p both to continue and to cease. The besirist will deny that the causal roles that are distinctive of besires are the same as the causal roles that are definitive of beliefs that are not desires and of desires that are not beliefs. Besires will share some but not all of the causal roles that are distinctive of beliefs, and some but not all of the causal roles distinctive of desires. (Zangwill 2008a footnote p. 51)

Suppose we think of belief and desire in terms of ‘direction of fit’, normatively conceived. The claim that there are no ‘besires’ then amounts to the claim that there are no mental states that ground *pro tanto* requirements to action that are also subject to *pro tanto* requirements based on how the world is. *Prima facie*, there is no good reason for accepting this claim, so if Humeans are to convince us they need to produce a good argument.

Smith has two other arguments against ‘besires’. The first is that, for any putative besire that p, there are cases of people who take it that p, but who are not motivated. Hence, any case of being motivated by a belief that p is best characterised as having a belief coupled with a distinct desire, rather than a besire. If it were a besire then it would be impossible to continue to take it that p whilst losing one’s motivation. However, on the belief-desire model it is easy to explain this: the agent has retained the belief but lost the associated desire. This argument does not work. Showing that some agents who are motivated by taking it that p have beliefs that are not intrinsically motivating for them does not show that all instances of taking it that p are not intrinsically motivating. Put another way, showing that on some occasions when someone takes it that p they are in a state of believing that p, not besiring that p, does not show that no one ever besires that p<sup>18</sup>.

---

<sup>18</sup> Perhaps this is what Zangwill is getting at when he asks, “why should we infer that some particular besire state is not a unitary state from the fact that there are some other states of beliefs without desires?” (2008a fn p. 51)

Smith's other argument is that there are no unitary states such as besires, because beliefs and desires "can always be pulled apart, at least modally" (Smith 1994 p. 119). If there were a besire that *p*, the belief-like part would go out of existence when not-*p* and the desire-like part would persist. Therefore the besire is not a unitary state: it can be pulled apart modally. But this reply will not do (Smith 1994 p. 118; cf. McDowell 1978; McNaughton 1988; Little 1997). The first problem is that it begs the question. If there are any besires then, necessarily, they cannot be pulled apart modally. Even if, for any belief and desire, it is possible for someone to have one and lack the other, this does not show that it is not possible for someone to be in a unitary state that is both cognitive and motivating. It does not show that there are no states that are motivating recognitions of requirements on action that are *also* accountable to norms external to the agent's psychological states. Being in the state of recognising a normative requirement would seem to be exactly a state of this kind.

The second problem is that such a response, even if it were not question-begging, would only rule out the possibility of a besire to *p* which is belief-like and desire-like with respect to the same proposition, *p*. Moral judgements of the form 'I am morally required to  $\Phi$ ' would not be like this. On the normative conception of 'direction of fit', such a judgement would be belief-like with respect to the proposition that  $\Phi$ -ing is my moral duty, and desire-like with respect to the proposition 'I am now  $\Phi$ -ing', 'I am about to  $\Phi$ ' or 'I have  $\Phi$ -ed'. That is to say, there is a *pro tanto* rational requirement that I withdraw this judgement upon my coming to believe anything which suggests that I am not required to  $\Phi$ . But while I am committed to this judgement, it grounds a *pro tanto* rational requirement that I  $\Phi$ .

In response to this possibility, Smith merely reiterates his assertion that beliefs and desires can always be pulled apart modally (1994 p. 119). However, we have seen no reason for insisting that a besire that is belief-like with respect to *p* and desire-like with respect to *q* can be pulled apart modally. We have not even ruled out that there might be some propositions such that if one is in a belief-like state with respect to them, one is necessarily in a desire-like state towards something else. Smith takes this to be what is at issue:

As I see it, the disagreement between Humeans and anti-Humeans as to whether we are just believers and desirers, or rather besirers as well, amounts to no more and no less than a disagreement about these modal claims. (1994 p. 120)



However, I do not think we should see things this way. The anti-Humean does not claim that there are some *propositions* such that, if one is in a belief-like state towards them, one is *necessarily* also in a desire-like state towards something else. The important point is that sometimes coming to take things to be a certain way – ‘believing’ something – can be motivating without the input of any independent desire. Such instances of being motivated by taking things to be a certain way are instances of what is called ‘besiring’<sup>19</sup>. This is not to say that moral judgements are always intrinsically motivating; in fact, it is not to say anything about whether they motivate. It is only to assert that they can ground *pro tanto* requirements on action. And this does not even commit us to the claim that rational agents necessarily act in accordance with their moral judgements, since they might sometimes (or often) rightly judge that other, non-moral, considerations should take precedence. So our rejection of the Humean theory of motivation does not commit us to any controversial claims.

## 6.4 Beliefs and desires, facts and values

The discussion of besires was not supposed to show that there are three, rather than two, fundamental kinds of mental state. Rather, it was meant to undermine the Humean claim that action is to be understood as the product of the two distinct kinds of mental states that might be called ‘cognitive’ and ‘motivational’ and classified under ‘reason’ and ‘the passions’. Thinking of beliefs and desires normatively already represents an important step in undermining this distinction. It allows us to see that there may be states of many kinds subject to different norms. We can think of *fear* and judgements of fearfulness, for example, as being subject to norms that show both their appropriateness as responses to fearful situations and their role in motivating appropriate (e.g. danger averting) actions. Freeing ourselves from the assumption that states must fall into only two kinds each bearing only a single asymmetrical relation to the world allows us to do justice to the complexity of experience, thought, motivation and action.

---

<sup>19</sup> Korsgaard says that practical reasons must be capable of motivating us (1996a p. 317), but whilst they do not necessarily move us, “their necessity may lie in the fact that, when they do move us... they move us with the force of necessity” (1996a p. 320). This may sound obscure. However, I think that Korsgaard means to say that the fact that a consideration does not motivate all agents does not mean that it cannot be intrinsically motivating for some agents. A consideration would be intrinsically motivating with the relevant sense of necessity if (a) no independent desire or motivational state is needed and (b) the consideration is of a kind that the agent finds they must acknowledge in their deliberations about what to do since they find that they cannot sincerely dismiss it as unimportant.

Once we begin to talk in this way, Humeans might come up with a new objection as follows: since (i.) the belief-desire distinction mirrors the fact-value distinction, (ii.) rejecting the former commits us to rejecting the latter, and (iii.) this is implausible. The first two parts of this claim (i., ii.) might be justified by saying that beliefs represent the world, and when they do so successfully they report *facts*. Our interests, projects and ends, on the other hand, not only form the content of our desires but explain why we find *value* in, or project values onto, the world<sup>20</sup>. The third part of the claim might be justified by saying that any attempt to reject the fact-value distinction is an instance of the naturalistic fallacy (Moore 1903 §§ 24-5) since it would allow, falsely, that normative or evaluative conclusions can be inferred from purely factual premises.

More would need to be said to clarify and support this claim, but let us accept this third claim for the sake of seeing whether the first two steps of the argument work. I claim that they do not, because respecting the fact-value distinction does not entail that mental states can be divided into belief-like and desire-like states without overlap or remainder. Respecting the fact-value distinction requires that we recognise that a valid inference to an evaluative conclusion must have an evaluative premise. However, it does not follow from this that the world can be neatly divided into evaluative and factual ‘parts’ or that the evaluative and the factual cannot be inseparably intertwined in aspects of human experience, thought, talk and practice. Consider the following example, given by Cooper:

[A] hill can only be disclosed as steep through an ‘attunement’ that accompanies activities like hiking. A steep hill is not one with a certain angle of elevation, but one experienced as daunting, whose conquest is cause for satisfaction and the like<sup>21</sup>. (1990 p.88)

The way we experience things, the ways we take the world to be, is not as a world of facts distinct from and independent of values and significance. The world is not a set of

---

<sup>20</sup> Or we might say that, just as in believing we aim to conform to facts, in desiring we aim to conform to values.

<sup>21</sup> One might wonder then what we should say about other instances of steepness, such as a steep curve on a graph. Does steepness here refer to having a particular angle of elevation, or is it a metaphor drawn from hill walking? There isn’t space to tackle the issue fully here, but it is worth noting that whether an elevation of line or curve on a graph is considered steep will often depend upon the significance of the data being plotted. A steep curve on a graph showing HIV infection is not one with a particular elevation (after all, one can change the labelling on the axes) but one showing an infection rate that is surprising, higher than expected or a cause for concern. The point is not that all properties must involve an

things onto which we *subsequently* project value through our desires or even through our normative commitments and evaluative beliefs<sup>22</sup>. As Bracken says, “human experience... is... always value-laden. We experience our world as, first and foremost, a world involving relationships of significance” (1999, p.84). Our experience is “of a world ready-to-hand, replete with significance, utility and worth” (Cooper 1990 p. 90).

Humeans might object to Cooper’s example, claiming that steepness is a subjective projection, not a genuinely mind-independent property. However, “[i]f ‘steep’ is to be described as ‘subjective’, so must the vocabulary of destruction, fragility, solidity and so on” (Cooper 1990 p.88). Humeans claim that we can have beliefs about steepness, solidity, fragility and destruction. If we can have beliefs about these properties and these properties are mere value projections onto the world, then the belief-desire distinction does not track the fact-value distinction. Our thought about the world picks out features that have particular significances for human life, purposes and projects. The fact-value distinction does not highlight two separable aspects of our experience; it is a (potentially useful) theoretical abstraction from our experience. Our experience of the world is not as mere information gatherers, as indiscriminate collectors of ‘facts’. It is purposive, embodied and involved, orientated around projects and people. Our beliefs and experiences are already practical in a certain sense. This sense of practicality is best illustrated by Heidegger’s famous example of hammering in a workshop. As Guignon puts it:

[W]hat we encounter when we are absorbed in such an activity is not a ‘hammer-thing’ with properties to which we then assign a use value. On the contrary, what shows up for us initially is the hammering, which is ‘in order to’ nail boards together, which is ‘for’ building a bookcase, which is ultimately ‘for the sake of’ being, say, a person with a neat study... [T]he light in the room, the workbench, the saw, the glue - all of these show up in their interconnected functionality in relation to our projects. (1993 p. 10, cf. Heidegger 1962 p. 98)

The practicality and significance of things in relation to our projects and purposes is basic, in that it indicates an irreducible aspect of all our lived experience. The kind of

---

intertwining of the factual and the evaluative, but that many of them do and there is no reason to insist that we can only understand things properly by pulling these aspects apart.

<sup>22</sup> Heidegger writes, “[I]n interpreting, we do not... throw ‘signification’ over some naked thing which is present-at-hand, we do not stick a value on it.” (1962 §150)

theorising in which we seek to think about things independently of human interests is secondary, an abstraction from lived experience<sup>23</sup>. As Cooper puts it:

If our Being-in-the-world is an embodied engagement with a world that ‘opens’ itself to us through our concerns and projects, there can be no reason to think that it will be disclosed only when we take stock and reflect. On the contrary, unless its features are revealed in a more ‘proximal’ way, there would be nothing to take stock of and reflect upon. If so, it must be wrong to suppose that reason is the faculty which discovers how the world is and passion merely the arena in which our subjective reactions to this discovery are played out. (1990 p.89)

It is for this reason that in our everyday moral thought and talk we use thick concepts, concepts in which evaluation and description are inextricably linked (Wiggins 1987). Williams (1985) gives ‘coward’, ‘lie’, ‘brutality’ and ‘gratitude’ as examples of thick concepts. Making a judgement that involves such a concept is a unified act that is both assessable in terms of truth and falsity and is, in a sense, practically, normatively and motivationally committing. A judgement that an act was cowardly can be true or false. But it is also a negative assessment of that act, with the implication that one would wish, and try, to avoid performing such acts oneself and think less of oneself if one did act in that way. This suggests that our judgements about the characteristics of things, about the way that things are, already involve practicality and relate to motivation and action. In saying this we respect the fact-value distinction since we do not conflate facts and values. However, in showing that there are phenomena that cannot be adequately accounted for if we insist upon trying to divide everything in fact and value ‘parts’, we note a limit to the usefulness of the fact-value distinction.

## 6.5 Cognitivist internalism about reasons judgements

We have already seen plenty of reasons to reject the Humean theory of motivation. There is a further one that returns to the issue which led us to discuss the Humean theory of motivation in the first place: judgement internalism. Suppose that moral judgements are made true by facts about the reasons that we have to act in some ways rather than others. If so then we might try to understand moral judgements, as Smith

---

<sup>23</sup> In science we do not get away from this, but rather systematically pursue particular projects and purposes through the use particular of tools and procedures. The fact-value distinction is useful for putting a certain amount of distance between science and ethics, but it does not thereby show science as neutral or the world as devoid of values. Only in a laboratory can we think like this. Only a creature with particular projects and values would be capable of or interested in adopting the perspective of impartial investigator.

suggests, in terms of beliefs about what we have reasons to do. If this is so, and rational agents are motivated by moral judgements, then it follows that either:

- (a) being rational involves having general desires such as the desire to do what is morally right (and externalism is true) or
- (b) rational agents who are motivated to act in accordance with their moral judgements are not, in such cases, motivated by an appropriately related desire and means-end belief

This is a dilemma that we have seen before: it prevented Smith from establishing cognitivist internalism, given that he accepts the Humean theory. Cognitivist internalism is incompatible with the Humean theory. This means that any good argument for cognitivist internalism would be an argument against the Humean theory. However, we don't yet have a good argument for cognitivist internalism. Nonetheless, it seems that the dilemma described needs only to be modified slightly in order to produce a powerful argument against the Humean theory.

We can argue along the following lines. Firstly, we can note that judgements about what (all things considered) one has most reason to do are judgements about matters of fact. Secondly, we can note the following thesis about practical rationality:

- 1 Practical rationality: Insofar as one is practically rational, one is motivated by what one believes there to be (normative) reason to do.

This entails the following claim:

- 2 When one judges that (all things considered) one has most reason to  $\Phi$ , one is either motivated to  $\Phi$  or one is practically irrational.

If one judges that overall, taken together, all the relevant considerations favour  $\Phi$ -ing, but one fails to even attempt to  $\Phi$  then one is irrational by one's own lights. Statements such as 'I knew it would be best overall to take the morning flight, but I took the evening flight' or 'I knew that, overall, all the relevant considerations favoured taking out medical insurance, but I decided not to' are paradigmatic examples of confessions of irrationality. In such cases, the question 'why?' is not a request for a justification of the

decision, not a request for more reasons. For if there were any more reasons to be mentioned then the judgement was not, after all, an overall judgement. Rather, to ask why someone acted contrary to what they took to be overall the best or most rational thing to do is to ask what happened to them to *cause* them to act so irrationally, what interfered with or disrupted their rationality. It is to ask for an *explanation*, not a justification.

On the basis of these considerations, we can conclude that:

- 3 There are at least some judgements about matters of fact, such as the judgement ‘that I have most reason to  $\Phi$ ’ that are intrinsically motivating for agents who are not practically irrational.

This is a kind of internalism: non-moral internalism, or internalism about reasons judgements<sup>24</sup>. Although it asserts nothing about *moral* judgements, it claims that there is an internal or necessary connection between motivation and judgements about what there is most reason to do. It seems that this form of internalism is true and yet inconsistent with the Humean theory of motivation, since it claims that beliefs of a particular kind are intrinsically motivating for some agents. Humeans may reply that all this shows is that such beliefs necessarily succeed in motivating rational agents, but *not* that they are intrinsically motivating. The motivation, Humeans will claim, comes from the combination of that belief with an independent desire. Humeans then owe us some explanation of the necessity of the connection between the belief and the motivation. Such an explanation is easy for Humeans to find: the necessary connection is only found in rational agents, so being rational must involve having a desire to do what one believes one has most reason to do. At this point the anti-Humean response is two-fold. Firstly, why should we believe that rationality involves having any such desire? Positing such a desire seems like an *ad hoc* device designed only to save the Humean theory. Secondly, it seems that, in positing a desire that is constitutive of rationality, Humeans effectively abandon their position by admitting that desires can be rationally required. It only makes sense to say that desires can be rationally required if they can be rationally acquired. And if desires can be rationally acquired then there is no reason to deny that reason can motivate.

---

<sup>24</sup> This is completely different from the internalism about reasons that is mentioned in section 6.7 (below) and discussed more fully in chapter seven.

## 6.6 Nihilism and Humeanism

So far we have only looked in detail at various forms of a single argument for the Humean theory of motivation – Smith’s argument. However, Shafer-Landau (2003) discusses four other arguments. I now look at these to see whether they provide the Humean theory with any support, and conclude that they do not. The first is “the *argument from moral nihilism*” (Shafer-Landau 2003 p. 127), attributed to Blackburn (1985). Suppose there is no moral reality (moral nihilism). Given that beliefs are representational states that are true when they correctly represent things as they are, and false otherwise, it follows from moral nihilism that moral beliefs are always false, because they have nothing to represent. If our moral judgements were beliefs it would then follow that they are uniformly in error (Mackie 1977). However, since our moral judgements are not uniformly in error, our moral judgements cannot be beliefs<sup>25</sup>.

Does this argument succeed? Firstly, moral nihilism is highly controversial, so even if the argument were conclusively shown to be valid, it would remain controversial as to whether it is sound<sup>26</sup>. Secondly, the argument is overhasty in dismissing error theory. The fact that it would be disastrous if our moral judgements were uniformly in error may provide us with a good reason to produce an argument that they are not, but it is not itself an argument. Thirdly, a positive argument is needed as well as a negative one; showing that moral judgements cannot be beliefs does not show that, for example, they are desires. An examination of the nature of moral judgements is required. Fourthly, even if moral nihilism is true and error theory false, the argument simply does not entail that the Humean theory of motivation is true. The fact that moral judgements are not beliefs does not entail that *had* moral judgements turned out to have been beliefs they would not have been capable of motivating rational agents. And even if it had entailed this, this would not entail that beliefs in general are not capable of motivating rational agents. It certainly does not entail that a belief-desire pair are necessary and sufficient for an agent to be motivated to act or provide a constitutive explanation of action. These deficiencies seem not to indicate that there some gaps in the argument

---

<sup>25</sup> Shafer-Landau is primarily concerned with whether or not this argument will ground, via the Humean theory, an argument against moral realism. Since in this context the argument from nihilism would be question-begging Shafer-Landau rejects it without considering whether it is valid (2003 pp. 127-8).

<sup>26</sup> i.e. as to whether the premises and conclusion are true.

from moral nihilism. Rather, they seem to indicate that the argument is actually irrelevant to the Humean theory of motivation.

However, there might yet be an interesting connection between the Humean theory and nihilism more broadly construed: nihilism not just about morality but about practical normativity. As Wallace points out, if there are no normative reasons, “no fact or truth about which considerations really do ‘recommend’ or ‘speak in favor’ of particular actions” (2006a p. 68), we still need to explain the existence of apparently reason-guided activity. If there are no normative reasons, this cannot involve pointing to anything in the world, so it must involve pointing out features of the agent’s psychology. Hence if there are no normative reasons, explanations of action must be Humean. As Wallace points out, if we are normative nihilists, “[w]hat renders the agent’s deliberative point of view intelligible must therefore be some noncognitive state of desire or pro-attitude, a kind of affect that is projected onto the world” (*ibid.*). So, Wallace claims, we misunderstand the Humean position if we think that Humeans are “realists about the normative who mistakenly think that normative reasons are constituted by desires” (2006a p. 69)<sup>27</sup>.

Wallace does not make it explicit how he thinks that normative nihilism entails the Humean theory<sup>28</sup>. However, it seems that nihilism does not even entail that action explanations must be psychologistic, and therefore cannot possibly entail that they must take the belief-desire form. The fact that there is nothing in the world that genuinely *favours* acting in a particular way does not mean that there is nothing in the world<sup>29</sup> that can *explain* acting in that way. However, nihilism does seem to push us the direction of the Humean theory. If there are no normative reasons then there are no reasons to want or value some things rather than others. If this were right then reason would seem to be motivationally inert – or at least, reason would seem to be motivationally alert when it is functioning as it should and we are seeing things aright. If we were confused and did not see the truth of nihilism, we might think that we have good reasons, this might lead us to action, and this might be the (mis)working of the faculty of reason. This is actually

---

<sup>27</sup> In considering a putative argument from internalism about reasons to the Humean theory I provide backing for this claim of Wallace’s. Internalism about reasons does not entail the Humean theory; the argument is more likely to run the other way, but clearly it can only do so if there is some good reason for accepting the Humean theory. This is why I have elected to discuss motivating reasons before normative reasons: we need to dislodge the Humean theory in order to avoid confusion on normative matters.

<sup>28</sup> Actually Wallace does not explicitly say that he thinks that normative nihilism entails the Humean theory. However, what he says strongly implies that normative nihilism, if true, would provide substantial support for the Humean theory.

<sup>29</sup> By ‘in the world’ here I just mean anything that is not a feature of the agent’s psychology.



quite consistent with the falsity of the Humean theory, since it describes a situation in which beliefs appear to motivate and does not give us any reason for thinking that desires or, more specifically, belief-desire pairs, must be involved. Therefore, the truth of normative nihilism would be insufficient to show the truth of the Humean theory. Moreover, anyone who argues for the Humean theory from normative nihilism forfeits their right to speak of motivating *reasons*, for strictly speaking, on the nihilist view, there is nothing reasonable or unreasonable about any action or any motivational state. We can conclude, therefore, that normative nihilism represents a serious threat to morality, but this is because it entails that there are no moral reasons and not because it supports the Humean theory of motivation. In any case, if my arguments in chapter eight are right then normative nihilism is false.

## 6.7 Three more Humean arguments

The second argument Shafer-Landau considers is “the *argument from internalism about reasons*” (2003 p. 128). The argument is as follows:

- (1) Internalism about reasons: “one has a reason to do something only if one has an appropriate desire (e.g. a desire to do that very action, or a desire for some end one believes achievable through that action)” (Shafer-Landau 2003 p. 128)

Therefore:

- (2) “correctly judging oneself to have a reason entails the existence of a desire” (*ibid.*)
- (3) Evaluative beliefs are not intrinsically motivating: they are extrinsically motivating, owing their motivation power to the independent desires that (are required to) make them true (given the truth of internalism about reasons (1) above).

Internalism about reasons is a controversial claim<sup>30</sup>. It is controversial both whether an agent’s normative reasons are dependent on their desires or subjective motivations, and if there is such a connection, exactly what its nature is. However, more importantly and as I now argue, the argument as presented above does not support the Humean theory since (3) does not follow from (2), and (3) does not even entail the Humean theory.

---

<sup>30</sup> If my arguments in chapter seven are right, it is also a false claim.

Shafer-Landau tries to show that (2) fails to entail (3) in the following way. He rightly points out that holding an evaluative belief is not necessarily the same thing as taking there to be a reason for action. Sometimes they coincide, but not always. He writes:

Even if an action's rightness, goodness, admirability, etc. in fact entail a reason for action, one may *judge* an action right, good, or admirable without judging that one has reason for action... one can judge something good or right without also believing that one has a reason for action, even if truths about goodness, rightness, etc. just are, or entail, truths about reasons for action (2003 p. 129)

Since the Humean argument under discussion is designed to show that an important class of counterexamples to the Humean claim are not intrinsically motivating, it takes as a datum the fact that evaluative beliefs at least sometimes appear to be intrinsically motivating. The Humean argument is that because the truth of such beliefs entails the presence of desires, they do not constitute a counterexample. But, Shafer-Landau responds, even if beliefs about reasons entail the presence of desires, evaluative beliefs do not, since someone might judge something morally right without thinking it gives them any reason for action<sup>31</sup>. However, in cases where someone makes an evaluative judgement without taking it to be reason-giving, it seems unlikely that they would be motivated by that judgement. And in cases where evaluative beliefs do not even appear to be intrinsically motivating they cannot constitute counterexamples to the Humean claim. So Shafer-Landau's response appears unsuccessful.

Nonetheless, (2) does not entail (3) even if we suppose that "evaluative beliefs *are* invariably beliefs about what reasons we have" (Shafer-Landau 2003 p. 129). If evaluative beliefs appear to be motivating, they appear motivating in cases when they are false just as in cases when they are true. Suppose we accept (1) internalism about reasons, we accept, that is, that the truth of a reason claim about an agent entails the existence of a desire of that agent. However, from the fact that the *truth* of a proposition, *p*, entails that an agent, *A*, has a desire, *d*, it does not follow that *A*'s *belief* that *p* entails that *A* has *d*. *A*'s belief might be false. In cases where a true evaluative belief motivates Humeans have a ready explanation: if evaluative beliefs are beliefs about reasons then the truth of the evaluative beliefs entails the presence of a desire. But

---

<sup>31</sup> I might even judge it morally right for me now to  $\Phi$  without judging that this gives me any reason to  $\Phi$  – if I am a moral sceptic. Even if moral scepticism is false and I am making a mistake in my judgement, it is still possible (if irrational or mistaken) for me to make the former judgement without making the latter.

in the case of a false evaluative belief, the assumption of internalism about reasons does not entail that the agent has a desire to do what they believe they should do since when an evaluative belief is false there might be no reason present. So if false evaluative beliefs can motivate then internalism about reasons does nothing to show that they are not a counterexample to the Humean theory.

There are another two objections to the argument from internalism about reasons. The first is that it misrepresents internalism about reasons by replacing it with a less plausible claim sometimes referred to as instrumentalism or ‘the sub-Humean model’ (Lillehammer 2000) of normative reasons. I don’t discuss this objection here, although instrumentalism is discussed in the next chapter. The second is that the argument from internalism about reasons may well beg the question against the anti-Humean, for it is very likely that the Humean theory of motivation is required as an assumption to establish internalism about reasons. If so, even if we accepted the argument from internalism about reasons, we would end up with a picture according to which the Humean theory and internalism about reasons entail each other, but we have no reason to accept either. Taken together, these considerations show that internalism about reasons cannot support the Humean theory. Internalism about reasons is discussed in detail in the next chapter.

Shafer-Landau discusses a third argument for the Humean theory that he calls “the *argument from identical beliefs*”:

1. If beliefs were sufficient to motivate, then two people with the same belief must be identically motivated
2. People with the same belief can be motivated differently.
3. Therefore beliefs are not sufficient to motivate. (2003 p. 129)

Shafer-Landau takes the argument to be valid and the second premise to be true, but the first premise to be false:

People with identical beliefs can be differently motivated because of other beliefs or desires that they don’t share. You and I both believe that a bus is hurtling toward us, that we are directly in its path and will be killed unless we move. I move and you don’t. We had different motivations, which prompted us to different actions. But these different motivations could be explained by different evaluative beliefs (I think it good to remain alive; you believe that life

is not worth living). Or they could be explained by different desires (I want to remain alive; you don't). (2003 pp. 129-30)

What Shafer-Landau's example shows is that in the situation in question, the belief 'that a bus is hurtling towards us' is *not* sufficient to motivate. What Shafer-Landau means to point out, and later does succeed in pointing out (although not in these terms), is that the argument conflates two sense of sufficient (discussed earlier in chapter three, § 3.5 fn. 6). In the first sense, to say that a belief is sufficient to motivate is to say that it is *intrinsically* motivating, and that we do not need to add anything to the belief in order for it to motivate. On the other hand, we could think of a belief that is sufficient to motivate as one that motivates no matter what other beliefs, desires or other mental states the agent has. This would be to think of being sufficient to motivate as being a guarantor of motivation. If a belief is sufficient to motivate in this sense, then nothing can remove or change its motivational force. The anti-Humean does not claim that beliefs are intrinsically motivating in *that* sense: perhaps there is nothing that is motivating in that sense. Desires, for example, are not guarantors of motivation. Two people can have the same desire but be differently motivated because of the other desires and evaluative beliefs they have (cf. Shafer-Landau 2003 p. 130). The motivational force of a belief or desire can not only be outweighed, it can be attenuated, defeated, enhanced, reversed or transferred to another action, behaving in much the same ways that normative reasons do on Dancy's (2004 pp. 38-52) account. To use a Humean type example, my belief that 'this train is going to Paris' might motivate me, together with my desire 'to get to Paris', to get onto the train. But if I come to believe that 'this train will take five hours to get there', and that 'the train on the other platform will only take two hours', and if I also desire 'to get to Paris in less than three hours', then this belief might not motivate me anymore. To give an anti-Humean example, the fact that it is raining might motivate me to, and give me a reason to, stand under a tree. If the rain shower develops into a thunderstorm, this might defeat this reason and motivate me to find somewhere safer to shelter. Given the distinctions between enablers, defeaters and motivators or motivational 'prompters', we can see why Shafer-Landau writes:

In many cases, an agent's strong desire not to  $\Phi$  explains why she doesn't  $\Phi$ . But this doesn't license us in attributing to an agent a desire whenever she  $\Phi$ s. That a desire can serve as a motivational defeater does not prove that it must be there as a motivational 'prompter'. (2003 p. 130)

Even if it is sometimes problematic to draw a distinction between the absence of a defeater and the presence of an enabler, this hardly helps Humeans. From the fact that a desire not to  $\Phi$  might explain not  $\Phi$ -ing, we can conclude *nothing whatsoever* about what explains the agent's  $\Phi$ -ing in the absence of such a desire. So the argument from identical beliefs is irrelevant to the Humean theory. It is based on the false assumption that in order for beliefs to be intrinsically motivating, they must be capable of overriding all other motivational factors.

The fourth argument for the Humean theory that Shafer-Landau discusses is the “*argument from continuity and uniformity*” (2003 p. 131):

[N]on-human animals are clearly the sorts of beings who can be motivated to do things. Yet such animals either lack beliefs entirely, or possess only instrumental, non-evaluative beliefs<sup>32</sup>. We ought to see ourselves as essentially continuous with the other inhabitants of the animal kingdom. If they don't need evaluative beliefs to get motivated, then there is a presumption that we don't need them, either. (2003 pp. 131-2)

So far the argument appears to be a complete non-sequitur: if it shows anything at all, it shows that evaluative beliefs are not *necessary* for motivation, but not that they cannot be sufficient or intrinsically motivating. However, note that Shafer-Landau distinguishes between different senses in which one can be an anti-Humean. Some anti-Humeans claim that (all) evaluative beliefs are sufficient to motivate. Other anti-Humeans add to this the claim that evaluative beliefs are *necessary* for motivation. In so doing, they will be “developing a completely uniform account of human motivation” (Shafer-Landau 2003 p. 134). According to Shafer-Landau, “others will espouse a hybrid theory that allows desires a primary motivational role in some circumstances” (*ibid.*), and this “is a sacrifice of theoretical uniformity” (*ibid.*). I don't see why we should attempt either of these tasks. A theory of motivating reasons is a theory that applies only to actions which are done for a reason. We can adopt a minimal conception of motivating reasons according to which a motivating reason is any consideration that the agent responded to in acting as they did, anything that was both taken to favour the action and that explains the action. In saying this we neither sacrifice ‘theoretical uniformity’ by offering a ‘hybrid theory’ nor claim that evaluative beliefs are necessary

---

<sup>32</sup> It is unclear whether this is true. If it is false it only serves to further undermine the Humean argument, so I assume for the sake of argument that it is true.

for motivation. Perhaps actions that are not done for a motivating reason are not motivated by evaluative beliefs, or perhaps responding to reasons does not have to involve having beliefs about those reasons.

To reject the Humean theory we do not need to argue that beliefs are necessary for motivation or that they are sufficient in the sense of guaranteeing motivation. The possibility we are trying to make room for is that beliefs can be intrinsically motivating or, better said, that our actions can be appropriate responses to our reflective appreciation of the way the world is. Shafer-Landau asks, “[w]hy talk of evaluative beliefs when all the relevant explanatory work can be done without them?” (2003 p. 132) He gets the answer to this challenge exactly right when he writes:

Our ability to take our lives and actions as objects of reflection, and to take that reflection as the basis of deliberation and action, signifies a relevant discontinuity from other animals. Yet this reflective ability may well receive expression in evaluative belief; indeed, that is precisely where we should expect to see its introduction (*ibid.*).

Belief-desire explanations obscure this essential feature of human actions. We can evaluate actions as better and worse, right and wrong, rational and irrational, and we are often moved by these evaluations to bring ourselves into line with how we judge we ought to be.

## 6.8 Conclusions

In previous chapters we saw that the Humean theory of motivation is not supported by commonsense, that it was not established by Hume and that it struggles to avoid becoming implausible without collapsing into a minimal anti-Humeanism. We also saw that Smith’s argument for the Humean theory was unconvincing and that there were good reasons for the thinking that the Humean theory is false. In this chapter we have seen that a normative conception of ‘direction of fit’ cannot rescue the Humean theory and that other normative assumptions render the Humean theory unstable so that it collapses into anti-Humeanism. We have also considered a number of other arguments for the Humean theory and concluded that none are successful. This being the case, we are justified in concluding that the Humean theory is false. Therefore, cognitivist internalism might be true. In the remaining chapters I discuss the nature of normative reasons, hoping to establish cognitivist internalism via the practical reason approach.

## Chapter 7. Normative Reasons and Subjective Motivations

### 7.0 Introduction

The Humean theory of motivation is false, so moral judgements might be intrinsically motivating. Agents *might* act because they recognise moral reasons. However, we have not yet shown that there are any moral reasons. According to instrumentalists and internalists about reasons, reasons for action are dependent upon subjective motivations. If this is right then, as I explain (§ 7.1), it casts doubt on the possibility of moral reasons. However, I argue that the arguments for instrumentalism and internalism about reasons are unconvincing and there are good reasons to reject both views. Then, in chapter eight, I argue that there are normative reasons for action, and their nature is such that it guarantees that moral reasons are counted among their number.

### 7.1 Why desire-based reasons are problematic for morality

According to instrumentalism and internalism about reasons, normative reasons are considerations that are appropriately related to subjective motivations that agents have. We can distinguish between three forms of instrumentalism. According to strong instrumentalism, I have a good reason,  $r$ , to  $\Phi$  if and only if I have a desire to  $\Psi$  and  $r$  is constituted by the fact that  $\Phi$ -ing subserves  $\Psi$ -ing. According to this thesis, having a desire is necessary and sufficient for having a reason<sup>1</sup>. Let us call the claim that having a desire is sufficient (but not necessary) for having a reason ‘moderate instrumentalism’ and the claim that having a desire is necessary (but not sufficient) for having a reason ‘weak instrumentalism’. ‘Internalism about reasons’ is the following claim:

I have a reason to  $\Phi$  if and only if there is some member,  $\Psi$ , of my subjective motivational set,  $S$ , such that if I were fully informed and deliberated rationally I would see that  $\Phi$  is a way to satisfy  $\Psi$ <sup>2</sup>.

However, any claim that having a reason is necessarily dependent upon having a desire<sup>3</sup> is problematic for morality because of the validity of the following inference<sup>4</sup>:

---

<sup>1</sup> Assuming there is actually some way the desire could be satisfied. This assumption is made throughout.

- (1) Morality can be justified only if it consists of categorical imperatives.
- (2) An imperative is categorical only if it gives a person reasons for action independently of what she desires.
- (3) Someone has reason to  $\Phi$  only if she desires to  $\Psi$  and (if she deliberated rationally she would see that)  $\Phi$ -ing is a way to  $\Psi$ .
- (4) Therefore any reason a person has depends on a desire of hers.
- (5) Therefore there are no categorical imperatives.
- (6) Therefore morality cannot be justified.

It may be good to have and act on moral desires. However, if one's *reason* for action is that the action would satisfy one's own desires, the action is necessarily not a moral one, even if the desire in question has moral content. 'That this action will satisfy my desire to help the poor' is not a moral reason, whereas 'that this action will help the poor' is. When one acts for the latter reason, one aims to help the poor. When one acts on the former reason, one's aim is the satisfaction of a desire and helping the poor is a mere means to gaining this satisfaction. The point is not that moral desires cannot play any role in explaining genuinely moral actions but that facts or beliefs about such desires are not moral reasons. Therefore, desire-based theories of normative reasons render morality problematic.

Desire-based accounts of reasons have another troubling consequence for morality. Internalism about reasons and all three forms of instrumentalism entail that, as Harman (1977 p. 107) infamously claimed, it is not true that Hitler ought not to have done what he did. Instrumentalism entails this by entailing one or both of the following claims<sup>5</sup>:

- (1) If Hitler desired that  $\Phi$  this gave him a good reason to do things that brought it about that  $\Phi$ .
- (2) If Hitler did not desire that  $\Phi$  then he had no good reason to do anything that brought it about that  $\Phi$ .

---

<sup>2</sup> Assuming that I recognise that I am motivated to  $\Psi$ .

<sup>3</sup> This includes strong and weak instrumentalism but not moderate instrumentalism.

<sup>4</sup> Adapted from Dreier (1997 p. 91).

<sup>5</sup> Moderate instrumentalism entails (1). Weak instrumentalism entails (2). Strong instrumentalism entails both (1) and (2).



Internalism about reasons entails both of the following claims:

- (3) If, after deliberating rationally from his current motivations, Hitler desired that  $\Phi$  then he had a reason to do whatever brought it about that  $\Phi$ .
- (4) If, after deliberating rationally from his current motivations, Hitler did not desire that  $\Phi$  then he had no reason to do anything that would bring it about that  $\Phi$ .

Hitler wanted that ‘Jews be taken to concentration camps’ and did not want that ‘use of concentration camps be discontinued’. Substituting the former proposition into 1 and 3 and the latter into 2 and 4 shows why, if either instrumentalism or internalism about reasons were true, morality would be in serious trouble.

## 7.2 Arguments for instrumentalism

Fehige accepts strong instrumentalism. He writes:

What fulfils our desires is good for us to get, what fulfils our desires is rational for us to do. Rationality is prudence. It is the intelligent pursuit, within the limits of the available information and resources, of our goals and projects, and thus of our own good. (2001 p. 60)

To establish instrumentalism in this manner one must show that whatever fulfils our desires is good for us to get. Fehige argues that “[o]ur notion of welfare must include a subjective element” (2001 p. 58) and, because “everyone is better off feeling good than feeling bad” (p. 59) part of this subjective element is pleasure and the absence of pain. However, Fehige claims pleasure cannot be the sole determinant of welfare, because otherwise it would be better to be connected against one’s will to a ‘pleasure machine’ for the rest of one’s life. Since this would not be better for one, there must a second constituent to welfare: desire fulfilment. Fehige says that since everyone desires their own pleasure, every instance of pleasure is an instance of desire fulfilment, and therefore our notion of welfare is the notion of desire fulfilment.

Fehige’s argument is unconvincing. Taking pleasure in someone else’s suffering is a way of feeling good, but is worse than feeling bad about their suffering. The observation that, *other things being equal*, people are better off feeling good than feeling bad, fails to show that feeling good is always better overall than feeling bad. Yet

Fehige wants us to draw the even stronger conclusion that feeling good is one of only two determinants of welfare. Fehige ignores the fact we can apparently desire and take pleasure in what is not good, even in what is not good *for us*. Indeed, we can desire things that are bad for us (desiring to smoke a cigarette is an example). This strong anti-instrumentalist intuition needs to be countered and explained by instrumentalists. Fehige no doubt wants to maintain that a desire to  $\Phi$  is always a contributory reason to pursue the means to  $\Phi$ , and that in cases where  $\Phi$ -ing is not conducive to one's welfare this must be because one has stronger desires that will be frustrated by this action. This response is perfectly valid once instrumentalism has been established, but until it has non-instrumentalists may quite reasonably ask why this *must* be the case: why *must* it be true that the agent's welfare always lines up with the agent's strongest desires?

There is a second problem for Fehige. Fehige's argument proceeds as follows:

[T]he person who believes an action to be best for him, but doesn't perform it, is irrational... In other words, it is rational to do what he believes is best for him. As we have just argued, 'best for him' means 'best fulfills his desires'. Thus, it is rational for him to do what he believes would best fulfill his desires. Ditto for each of us. (2001 pp. 59-60)

This argument does not succeed. Firstly, the following is an open question: ' $\Phi$ -ing best fulfils my desires, but is it best for me?' Therefore 'best for me' does not *mean* 'best fulfils my desires'. Secondly, many agents in fact reject instrumentalism. For such agents believing ' $\Phi$ -ing would be best for me' is compatible with believing ' $\Phi$ -ing does not best fulfil my desires'. Failing to do what one believes best may be irrational, but it is unclear whether there is any such irrationality in not doing what one believes best fulfil one's desires. Thirdly, the phrase 'a person who believes an action to be best for him' admits of two interpretations. An agent, *A*, might believe that, of the actions that *A* can perform, there is overall most reason to  $\Phi$ . Alternatively, *A* might believe that, among the actions that *A* could perform,  $\Phi$  is best *for A*. Only on the first interpretation does it follow without additional argumentation that the action that is best for him is rationally required. But only on the second interpretation is it necessarily the case that what best fulfils his desires is best for him. Therefore, we have no reason to believe that 'best fulfils his desires' entails 'is rationally required'.

Two more ambiguities, once noted, undermine Fehige's argument. 'Rational' admits of various interpretations, including 'rationally required' and 'rationally intelligible'. When an agent does what best fulfils their desires their action is rationally

*intelligible*, but it does not follow that their action was rationally required. The instrumentalist insistence that we should draw this stronger conclusion is unjustified. ‘Best’ is also ambiguous. It might mean ‘most effectively’. On this reading that which best fulfils our desires fulfils the *greatest number* of those desires, or fulfils them to the *greatest extent*. However, ‘best’ could have a non-maximising interpretation: that which fulfils the most appropriate of our desires or fulfils our desires most appropriately. Alternatively, it could mean that which fulfils our desires in the most *moral* way. If we use ‘best’ in a maximising sense it seems implausible that we are rationally required to do what best fulfils our desires. However, only the maximising sense of ‘best’ can yield the instrumentalist conclusion.

Fehige’s argument fails to show that desires have special normative status: that they, and the means to their satisfaction, are the *only* considerations that provide reasons for action. Dreier also attempts to argue for this claim by arguing that the *only* principle of practical reason that is categorically binding is the means-end (M/E) principle:

M/E: If you desire to  $\Psi$ , and believe that by  $\Phi$ -ing you will  $\Psi$ , then you ought to  $\Phi$ . (Dreier 1997 p. 93)

Dreier imagines trying to persuade someone, Ann, that she ought to  $\Phi$ . Citing a rule,  $R$ , directing her to  $\Phi$  is only effective if she accepts  $R$ . Dreier accepts the Humean theory of motivation, claiming that Ann is motivated to  $\Phi$  only if she has a desire to  $\Psi$  and a belief that by  $\Phi$ -ing she would  $\Psi$ . Therefore, according to Dreier, Ann is motivated to  $\Phi$  by  $R$  only if she believes that  $\Phi$ -ing is a way to conform to  $R$  and desires to conform to  $R$ . Dreier claims M/E is the *only* categorical principle of practical reason for the following reason. Suppose someone desires to  $\Psi$  and believes that by  $\Phi$ -ing they will  $\Psi$ , but refuses to accept M/E and therefore refuses to accept that they ought to  $\Phi$ . What, Dreier asks, is this person missing? Adding a desire to comply with M/E will not help. This would only entail ‘I desire to comply with M/E and believe that by  $\Phi$ -ing I would comply with M/E’. The motivating force of this belief-desire pair is itself dependent on conformance to M/E and it is M/E that the agent refuses to accept. Dreier says, “[o]nce you have (accept) the means/ends rule, what you need to get you to acceptance of other rules is one or another desire. But no desire will get you to the means/ends rule itself” (1997 p. 94). Dreier claims that since M/E specifies what it is for something to be a normative reason, we do not need a normative reason to follow M/E, since to ask for a

normative reason is already to assume M/E. Therefore M/E is categorical in a special sense.

There are three reasons to reject Dreier's argument<sup>6</sup>. Firstly, Dreier's overall strategy is flawed. He assumes M/E is normative and then asks whether it is categorical or hypothetical. Having shown that it cannot be hypothetical, he concludes it is categorical. This ignores the possibility that M/E is not a normative principle of practical reason at all. Showing that one *can* comply with a principle does not show that the principle is *normative*. Secondly, Dreier's argument only establishes that *if* one can comply with M/E then one can do so without having a desire to do so. It does not establish that one *can* comply with M/E. Thirdly, Dreier's argument is dependent on the Humean theory of motivation. Since the Humean theory is false one may be able to comply with *any* principle without having a desire to do so. Therefore, M/E has no special status. If Dreier's argument had succeeded then perhaps it would have shown that we have reason to conform to M/E, namely that without so conforming we cannot have any other practical reasons. However, since the Humean theory is false we lack this reason for conforming to M/E, since conformance to other norms is possible independently of any desire to do so and hence independently of conformance to M/E. Therefore Dreier's argument does not succeed in showing that M/E is normative and provides no support to instrumentalism.

### 7.3 Arguments against instrumentalism

According to instrumentalism whatever I most desire to do is what I have most reason to do (strong instrumentalism) or whatever I have some desire to do I have some reason to do (moderate instrumentalism) or whatever I have some reason to do I have some desire to do (weak instrumentalism).

However, motivations and reasons – even recognised reasons – can come apart. Kleptomaniacs, in stealing despite resolving not to (Smith 1994 p. 133; Ayer 1954 p. 20), are motivated to do what they know they have no reason to do. Similarly, “a woman who has a sudden urge to drown her bawling baby in the bath” (Watson 1975 p. 101) “acts without thinking that what she does is rationally justified at all” (Smith 1994

---

<sup>6</sup> It might be thought that two further facts count against instrumentalism: that desires can conflict and that when we desire to  $\Phi$  we can have second order desire not to  $\Phi$ . Instrumentalists can counter the first threat by simply pointing out that there can be reasons both for and against performing an action. Dealing with the second issue appears more complicated.

p. 134). Despite momentarily wanting to drown her baby, she rightly recognises that there is not even a weak, defeated, reason for her to do so. These situations show that having a desire, even in conjunction with a belief about how to satisfy that desire, is insufficient to provide a good reason for action. Just as desires may fail to give us reasons, we may have reasons without desires. As Johnson says, “often people are not motivated in the least by values, even those they embrace” (1999 p. 53). It is possible (although not rational) for us to acknowledge that there is a reason for action yet remain unmotivated. If reasons were based in desires this would be impossible. These observations undermine all three forms of instrumentalism.

A further reason for rejecting *any* account that bases normative reasons in desires comes from Nagel’s distinction between motivated and unmotivated desires. Motivated desires “are arrived at by decision and after deliberation” (Nagel 1970 p. 29). They admit of rational or motivational explanation (Schueler 1995 p. 16). Unmotivated desires do not, although they admit of causal explanation. Nagel writes:

Hunger is produced by lack of food, but is not motivated thereby. A desire to shop for groceries, after discovering nothing appetizing in the refrigerator, is on the other hand motivated by hunger. Rational or motivational explanation is just as much in order for that desire as for the action itself. (1970 p. 29)

Motivated desires are reason-based whereas unmotivated desires are not. This explains why desiring something often coincides<sup>7</sup> with having a reason to get it, but having a desire *by itself* does not give us a reason for action. Generally, we only have reason to want what we have a reason to get, and reason to want to do what we have reason to do. Therefore, one who has a motivated desire usually has a reason for action<sup>8</sup>, but whilst the desire is indicative of the reason, the reason for action is not provided by the desire itself. The desire is based on the reason.

Can unmotivated desires provide reasons? Consider hunger. Hunger is a bodily feeling that typically grounds a desire to eat<sup>9</sup>. However, this desire itself does not ground a reason. Hunger grounds reasons to eat of at least two kinds. Firstly, that eating

---

<sup>7</sup> In some cases one can have a reason for desire without a reason for action. I might have good reason to want it to rain tomorrow, but since there is nothing I can do to bring about rain tomorrow this gives me no reason for action. In other situations, I may be able to bring about that which I have a reason to want, but it may be inappropriate for me to do so. I may have a reason to want my friend’s work permit application to be processed quickly and if I work in the office where such applications are processed I may be able to bring this about. However, I might still lack any reason to bring it about.

<sup>8</sup> Assuming the motivated desire is based on a good reason.

would provide satisfaction is often a reason for eating provided by hunger. Secondly, hunger indicates that the body needs food, and ‘that the body needs food’ is a reason to eat. Neither of these reasons is provided by the desire to eat. We can imagine someone being given a drug that maintained their hunger so that it could not be satisfied. The desire to eat would remain. However, neither this desire nor this hunger grounds any reason to eat because (a) neither the desire nor the hunger indicates that the body needs food and (b) there is no prospect of satisfaction<sup>10</sup>.

This picture of the relation between desires and reasons is far more plausible than the instrumentalist alternative. However, there are two apparent counterexamples to it: urges and inclinations. Urges “which the person concerned entirely disavows” (Dancy 2000a p. 36) are not based on normative reasons. However, it is implausible that they provide reasons since disavowing an urge involves taking oneself to have reason to rid oneself of that urge and not pursue it. Urges that are not disavowed are urges that one to some extent endorses, and hence that one takes oneself to have some reason to pursue. Hence, they are based on reasons, rather than reasons being based on them. Of the other kind of putative counterexample, inclinations, Dancy writes:

Perhaps I wear this shirt today because I feel like it. Is this a case of a desire that is not based on a reason? No. There are perfectly good reasons for putting on this shirt. But they are not decisive, in the sense that there are equally good reasons for putting on a different one. Where the reasons are not decisive, I let inclination decide. But what I do, I have perfectly good reasons for doing, and I do it for those reasons. Inclinations are just desires that we have for inconclusive reasons (*ibid.*).

Inclinations are no counterexample to the claim that desires are based on reasons. Therefore, the reason-based-desires remains more plausible than the instrumentalist alternative.

There are other reasons for rejecting instrumentalism. Firstly, “there are adverse conditions, such as massive confusion, in which desires are irrationally possessed or acquired, thereby failing to give reasons for action” (Lillehammer 2000 p. 157). Secondly, instrumentalism entails, falsely, that “I may have no reason to prepare for

---

<sup>9</sup> Some cases of hunger do not ground such a desire. For example, extreme hunger sometimes involves nausea and aversion to eating.

<sup>10</sup> A third reason for eating is the intrinsic enjoyment of it. This persists in the drug example. It is also arguably dependent not only on the presence of (at least a moderate degree of) hunger but also on there being at least some desire to eat. However, the desire here is still neither the reason nor the ground for the reason. The ground for the reason is the prospect of pleasure.

what I know I shall have reason to do tomorrow” (Nagel 1970 p. 40). Thirdly, vice can render reasons invisible to us and involve our lacking desires that we later recognise we had reason to have. The example of Ebenezer Scrooge in *A Christmas Carol* illustrates this point (Goldstein 2004 p. 67); his insensitivity meant that he failed to care even about the fact that his miserliness and uncharitable nature led to his misery. Agents may have vices or deficiencies that do not affect their reasoning and deliberative capacities but which nevertheless prevent them from appreciating certain reasons (Goldstein 2004 p. 80).

Desires do not, by their mere presence, provide reasons or indicate the presence of reasons. As Dancy puts it:

[I]f someone wants something that there is no reason to want, his desire does not give him some reason for doing it, a reason that was not there before. If an action is silly, wanting to do it does not make it any the less silly (2000a pp. 36-7).

If, when standing on the top of a cliff, I feel a sudden desire to throw myself over the edge this does not provide me with any reason to do so<sup>11</sup>. Nor does it follow that I have any reason to do anything – such as climbing over the fence that separates me from the cliff edge – that is a means to that end<sup>12</sup>.

## 7.4 Williams’ argument for internalism

Bernard Williams (1979) distinguished between internal and external reasons<sup>13</sup>. An agent’s reason is internal only if it is suitably related to a motivation that the agent would have after informed rational deliberation. External reasons display no such dependence. Williams claims there cannot be any external reasons. I argue that, on the

---

<sup>11</sup> It might be thought that if I do genuinely desire to throw myself off a cliff then I do have a subjective reason to do so, but no good or objective reason to do so. However, this thought is not compatible with instrumentalism, which claims that all good reasons, being based on desires, are subjective reasons. Moreover, if subjective reasons are not good reasons then they are not normative reasons at all – but instrumentalism is supposed to be a thesis about good or normative reasons, not merely about motivations.

<sup>12</sup> Suppose that, as Audi suggests (1996 p. 400), I have no desire to avoid pain and death, perhaps because I have been manipulated. I still have a reason not to take a poison that will lead to a painful death. Instrumentalism, falsely, entails otherwise.

<sup>13</sup> See Parfit (1997) for a more comprehensive survey of the varieties of internalism.

contrary, all reasons for action are external<sup>14</sup>. Williams' internalism about reasons is the following claim:

All normative reasons are internal. An agent has an internal reason to perform an action if and only if, were the agent to know the relevant facts and rationally deliberate they would be motivated to perform that action<sup>15</sup>.

Williams argues as follows. He defines 'S' as the agent's "subjective motivational set" (1979 p. 18), which contains the agent's desires, as well as "dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agent" (p. 20). Williams says, "if something can be a reason for action, then it could be someone's reason for acting on a particular occasion, and it would then figure in an explanation of that action" (p. 22). The truth of an external reason statement about someone, Owen, is insufficient to explain their action,  $\Phi$ , because Owen can have an external reason to  $\Phi$  even when he has no motivation to  $\Phi$ . Believing oneself to have a reason typically motivates one to act. When Owen does believe the external reason statement it motivates him to  $\Phi$ . Then he has a motive in his S that is served by his  $\Phi$ -ing: the motive to  $\Phi$  itself, which is provided by the external reason statement. Now the reason is *internal*, not external. External reasons are distinctive because in *coming to believe* that one has an external reason to  $\Phi$ , one moves from being in a state where one has no motive that would be served by  $\Phi$ -ing to having a motive to  $\Phi$  that is provided by this belief.

Williams supposes externalists must maintain "that the agent should acquire the motivation *because* he comes to believe the reason statement, and that he should do the latter, moreover, because, in some way, he is considering the matter aright" (1979 p. 24)<sup>16</sup>. He continues:

If the theorist is to hold on to these conditions, he will, I think have to make the condition under which the agent appropriately comes to have the motivation

---

<sup>14</sup> I do not claim that no reason can be related to an agent's motivations but that reasons are not such that their status as reasons is dependent upon being suitably related to a motivation that the agent would have after informed rational deliberation.

<sup>15</sup> See, for example, Parfit 1997 p. 100. It is sometimes claimed that an agent also has an internal reason if the action would help to fulfil one of the agent's present desires. However, internalism is more plausible if this disjunct is removed so that it is claimed that only desires that would survive informed rational deliberation are reason-giving.

<sup>16</sup> The latter claim seems false. Externalists can allow that mistakenly believing oneself to have an external reason can explain action.



something like this, that he should deliberate correctly; and the external reasons statement itself might be taken as roughly equivalent to, or at least as entailing, the claim that if the agent rationally deliberated, then, whatever motivations he originally had, he would come to be motivated to  $\Phi$ . But if this is correct... all external reasons are false. For, *ex hypothesi*, there is no motivation for the agent to deliberate *from*, to reach this new motivation (*ibid.*).

Williams makes two unjustified assumptions. Firstly, he assumes that all deliberation must begin from an agent's motivations. He provides no argument for this claim. Secondly, Williams supposes that reasons are dependent on counterfactual overall motivations. This conflates contributory reasons with overall judgements. It is possible that I may have some reason to do one thing, more reason to do something else, and most reason to do a third thing (see Dancy 2004 ch. 2). Williams' account entails, falsely, that I can never have more than one reason to perform an action and can never have outweighed reasons against performing the action that I have most reason to perform. This mistake is caused by thinking of reasons as the outputs of (counterfactual acts of) deliberation, rather than the points from which deliberation begins<sup>17</sup>. This in itself is reason enough to reject internalism.

## 7.5 The anticipated externalist objection

Williams anticipates the following objection from externalists:

It might be said that the force of an external reason statement... implies that a rational agent would be motivated to act appropriately... because a rational agent is precisely one who has a general disposition in his *S* to do what (he believes) there is reason for him to do (1979 p. 24).

Williams responds:

[T]his reply merely puts off the problem. It reapplies the desire and belief model (roughly speaking) of explanation to the actions in question, but using a desire and a belief the content of which are in question. What is it that one comes to believe when he comes to believe that there is reason for him to  $\Phi$ , if it is not the proposition, or something that entails the proposition, that if he deliberated rationally, he would be motivated to act appropriately? (*ibid.*)

---

<sup>17</sup> When we deliberate we take into account considerations that we take to genuinely favour or require acting in a particular way. We start from contributory reasons and progress towards an overall judgement about what there is most reason to do.

Given my criticisms of Williams' original argument, it should be clear that Williams is not entitled to merely assume that what we have reason to do is determined by what we would be motivated to do after rationally deliberating from our pre-existing motivations. However, another problem for Williams is clear here as well: this brings out explicitly his reliance on the Humean theory of motivation. Williams assumes that externalists share this assumption, when externalists may in fact reject internalism precisely because they reject the Humean theory.

Appealing to the Humean theory appears to be the *only* way to rule out the possibility of external reasons and to support the assumptions that lead to internalism. We can see internalists as offering the following argument:

1. It is a conceptual truth that any reason statement must motivate any rational agent who comes to believe it.

Therefore:

2. If a rational agent comes to believe a statement that predicates something of an action but is not motivated by it then the statement cannot be a true reason statement.

However:

3. Beliefs are not intrinsically motivating.

Therefore:

4. Reason statements are made true by standing in a rational relation to the existing subjective motivations of rational agents.

I think that something like this does underlie Williams' argument, since he mentions Hume at crucial points in his argument (1979 pp. 23-4) and his position is produced by "addition and revision" to a 'sub-Humean' model (1979 p. 18). However, if it is possible for beliefs to be intrinsically motivating then this argument collapses. Since the Humean theory of motivation is false it *is* possible for beliefs to be intrinsically motivating<sup>18</sup>. Therefore, the argument collapses. There *might* be external reasons, so internalism is false.

---

<sup>18</sup> Williams seems to accept that beliefs can be intrinsically motivating, since he claims it is highly plausible that "believing that a particular consideration is a reason to act in a particular way [can] provide, or indeed constitute, a motivation to act" (1979 p. 23). However, he does not take this phenomenon seriously enough. If the Humean theory is false then we have no reason not to take this phenomenon at face value, and this means that the internalist argument fails and that internalism is false.

Suppose that someone is lying comatose in a doorway. This might be a reason for me to call an ambulance. According to anti-Humeans, my noticing that someone is lying comatose in a doorway or my realisation that I should call an ambulance could *itself* explain my calling an ambulance. Therefore this reason can motivate independently of any prior motivations, I can acquire the motivation *because* I came to believe the reason statement and in so doing see the matter aright. Anti-Humeanism entails that external reasons are possible. Therefore, *any* argument for internalism must presuppose the Humean theory. Since the Humean theory is false, there is no good argument for internalism. Since anti-Humeanism is true, external reasons are possible.

There are a couple of other things to note about Williams' response to the anticipated externalist objection. Firstly, Williams demands to know what it is that we believe when we believe that we have reason to  $\Phi$ . If we reject the internalist account of normative reasons, we need to provide an externalist one in its place. I consider this question in chapter eight. Secondly, we should note that the objection that Williams puts into the externalist's mouth is not actually externalist<sup>19</sup>. He imagines that externalists need to posit a general desire to do what there is reason to do in order to explain how external reasons can motivate. However, externalists claim that reasons are not dependent on *any* desire that the agent has prior to deliberation, not even a general desire to do what there is reason to do.

## 7.6 McDowell's externalism

There is another reason for rejecting Williams' assumptions. Recognition of external reasons may be capable of motivating agents even if there is no process of *reasoning* that could bring them to this point. McDowell suggests that in such cases, "the transition to being so motivated is a transition *to* deliberating correctly, not one effected *by* deliberating correctly; effecting the transition may need some non-rational alteration like conversion" (1995 p. 78). Failing to recognise a reason is not necessarily indicative of irrationality<sup>20</sup>. Rational deliberation may be insufficient to bring us to recognise the reasons we have<sup>21</sup>. As McDowell says, "[i]t would be odd to say that a person who finds

<sup>19</sup> It is an expression of what Wallace calls 'meta-internalism' (1999). Many of the arguments against internalism apply equally to meta-internalism. Meta-internalism presupposes the Humean theory, for instance. This is reason enough for us to reject it. Wallace provides other compelling reasons for rejecting meta-internalism (1999 pp. 227-36).

<sup>20</sup> Although failing to be motivated by a reason one does recognise *is*.

<sup>21</sup> We should therefore reject the way Smith (1994 p. 166) and Cowley (2005 p. 351) characterise the debate. They both think that externalists claim that rationality guarantees convergence in desires.

no reasons to listen to twelve-tone music is irrational, even though one thinks that the reasons are there” (*ibid.*).

Williams assumes that the purpose of externalism is to facilitate accusations of irrationality against agents whenever they refuse to accept external reason statements (1979 p. 25). Both Williams and Cowley point out that merely repeating an external reason statement is “mere ‘bluff’ and attempted browbeating within a context of a power struggle” (Cowley 2005 p. 351 cf. Williams 1995 p. 40). However, this does nothing to show that there *aren’t* any external reasons. It suggests that what is claimed to be an external reason is sometimes nothing of the sort. Regardless of whether a purported reason is a genuine reason or not, if the other person does not see this then *mere insistence* on this point amounts to manipulation. However, this applies to internal reasons just as much as to external reasons. What this shows us is *not* that reason claims must be rationally justified by showing that they are the results of sound informed deliberation from the agent’s desires, but merely that they must be rationally justified.

Cowley denies that McDowell’s insight establishes an advantage for externalism. Cowley discusses the example of Owen Wingrave who, despite his grandfather, Sir Philip, insisting that he should join the army, sees himself as having no reason to do so. Cowley, imagining a situation in which Owen admits that his grandfather had been ‘right all along’ about the reason that he had to join the army, writes:

I suggest that Owen ... changes his mind rather than comes to recognise some external truth. Despite his regretful words about how he ‘wished’ and ‘should have’, he also knows that it was impossible at  $t_1$  – given his determinate psychological state at  $t_1$  – for him to accept R. His regret is not that he failed to accept a reason he now accepts, his regret is that the course of action suggested by R was not somehow forced upon him by circumstances and against his considered judgement at  $t_1$  (2005 p. 364).

Cowley then asks us to imagine that Owen changes his mind again after “he goes off to battle and finds out how nasty it all is” (*ibid.*). Externalists may try to maintain that this time he has really discovered what he had reason to do all along. Cowley then asks: “[h]ow can the externalist be confident that this discovery is somehow final?” (2005 p. 364)

---

However, being rational is insufficient to guarantee that we will recognise all the reasons for action that we have. This is true unless being rational is supposed to involve perfecting every kind of practical and intellectual virtue.

None of this rules out the possibility offered by McDowell. Moreover, we do not have to accept the picture Cowley offers. Firstly, Cowley's account of Owen's regret is not compelling. After Owen changes his mind and decides to join the army, we *can* potentially both see him as recognising a reason that was 'there all along' *and* as being glad that he had not given in to his grandfather's attempts to get him to join the army – glad that he had not *pretended* to recognise a reason that he could not see. Secondly, we need to distinguish between contributory reasons and overall judgements. When Owen changes his mind for the second time, upon finding out how nasty battle is, we can see him not as changing his mind about the original reason but as discovering a new, additional (contrary) reason. Hence, his overall judgement changes, based, as it is, on the contributory reasons that he is aware of. It is not part of externalism to be confident that one's overall judgements always get things right or cannot be improved upon. Since our judgements are made only on the best information available at the time they are often tentative, requiring ongoing revision.

Cowley also discusses Owen's reflections on Sir Philip's reason for enlisting. The considerations that failed to convince Owen did not fail to convince Sir Philip; they explain his enlisting. Cowley concludes that therefore "Owen needs nothing further by way of explanation or justification" (2005 p. 352). However, this seems plainly false. Owen takes it that Sir Philip's enlistment was *not* justified. This is because, according to Cowley, "Owen is confident in his judgement that war is 'crass barbarism' and that Coyle's hero Napoleon is 'a scoundrel, a criminal, a monster'" (2005 p. 351, cf. James 1992 p. 51). If Owen is confident in these judgements, he takes it that there are good reasons for *anyone* not to enlist, for these judgements are in no way contingent upon the subjective motivations of either himself or Sir Philip.

## 7.7 Smith's objection to internalism

Internalism entails that normative reasons are agent-relative since it entails that normative reasons are relative to motivational sets and different agents have different motivational sets. Williams writes:

If an agent really is uninterested in pursuing what he needs; and this is not the product of false belief; and he could not reach any such motive from motives he has by the kind of deliberative processes we have discussed; then I think we do that to say that in the internal sense he indeed has no reason to pursue these things (1979 p. 21).

Smith argues that our concept of a reason is not relative. He points out that if it were relative then “we should expect to find that we are sometimes able to dissolve apparent disagreements by finding that both parties are speaking truly” (1994 p. 168). But this does not happen<sup>22</sup>. Moreover, Smith says, internalism undermines the normative significance of reasons. What I have internal reason to do is determined by processes of rational deliberation starting from my actual desires. Suppose after this process I end up with a desire that p. Does this give me a good reason to bring it about that p? Since according to internalists, my *actual* desires are not based on reasons<sup>23</sup>, they are contingent and normatively arbitrary, arbitrary from the point of view of reason or rationality. I might have had a completely different set of desires to start with, even, as Smith says, “a set that delivered up the desire that not p after a process of systematic justification” (1994 p. 172). Therefore it is arbitrary that after deliberation I desire that p rather than that not p. Therefore, the fact that ‘that p’ is the outcome of my reflections about my counterfactual desires cannot be normative, since “arbitrariness is precisely a feature of a consideration that tends to undermine any normative significance it might initially appear to have” (Smith 1994 pp. 172-3)<sup>24</sup>.

In response to Smith, Cowley distinguishes between “two kinds of objectivity” (2005 p. 357). In one sense, I experience a reason as objective when I experience it as distinct from my preferences. However, in the other sense a reason is only objective if it is “part of a *singular* objective realm of reasons, normatively binding on all” (*ibid.*). Cowley thinks that since reasons may be internal yet objective in the first sense internalism does not render reasons statements arbitrary. However, this response fails to connect with Smith’s original objection. I might experience a consideration as distinct from my preferences, see its connection with my ‘S’, and yet legitimately wonder whether it is a reason, whether that member of my S is one that I have any reason to endorse. If the members of the actual S are rationally contingent then, for any member of S with the content ‘that p’ we might equally have had one with the content ‘that not

---

<sup>22</sup> Reason claims can be relative in the sense that a parent has special obligations to their child that a stranger does not. However, this kind of relativity is not desire-based. It is grounded in obligations dependent on relationships. It is not based on the recognition of those obligations. A parent who neglects their child still has reasons to care for their child even if they care nothing about fulfilling their parental obligations. At any rate, this is what our prereflective intuitions tell us. Internalists tell us otherwise.

<sup>23</sup> This must claim this otherwise their account would be viciously circular.

<sup>24</sup> Cf. Smith (1989); Darwall, Gibbard and Railton (1992)

p'. Why should we seek to bring it about that p instead of adopting a desire 'that not p' into our S and seeking to satisfy that?<sup>25</sup> The arbitrariness worry persists.

Smith identifies another problem for internalism. As Cowley puts it, Smith claims that:

The very nature of many 'regulative concepts' must be non-relative as part of their regulative function. Williams seems to be allowing a massive contingency into all attempts at communication, which would make it miraculous for reasons to 'connect' at all. And yet our complex society would not be able to function without the vast majority of reason-based communicative attempts being successful. (Cowley 2005 p. 354)

Cowley counters the claim that internalism makes discussions about reasons instances of talking past one another by suggesting that "a great many reasons will be jointly recognised between individuals in virtue of having grown up in the same society" (2005 p. 358). If the reasons of two people are relative to their S but similarity in their S is guaranteed by having grown up in the same society then many of their reasons will be shared. Even if this is true, it does not follow that, when two people from the same society are talking about the reasons for action that they have, they are talking about the same thing. They are talking about two different things that happen, for the most part, to coincide. They are still talking past each other, albeit in a way that has all the appearances of successful communication. This has important implications for cases of moral disagreement where reasons are offered for competing moral claims. It entails that such disputants are talking past one another. It also entails that we can say nothing about what is to be done in such cases of disagreement. If anything which is a reason for me could turn out not to be a reason for my disputant then this includes reasons to resolve disagreements in one way rather than another. Cowley counters by challenging externalists:

*Who* is to determine which reasons are correctly attributable to the singular realm of normative binding reasons? Who has access to Thomas Nagel's 'View

---

<sup>25</sup> There are two ways internalists might respond. The first, suggesting that choice of members of S must be arbitrary, is unsatisfactory. This is not only false but also fails to address the problem. The other option is coherentism. Coherentism about desires has the unfortunate consequence of recommending that people with evil desires become more evil. It also struggles to escape the arbitrariness worry. It seems difficult to rule out that the possibility that there may be very many coherent sets of desires, each generating their own set of reasons. It seems arbitrary to say that agents are given reasons by just one of these sets, even if it is the one that most closely resembles their actual desires. Why should it be that coherent sets of *desires* of *counterfactual* agents generate normative requirements on *actions* of *actual* agents?

from Nowhere', whence to assess whether this reason is in fact the right one? (2005 p. 358)

This is somewhat unfair to externalists. Externalism is only the negative claim that reasons are *not* dependent on subjective motivations: why should this require a 'view from nowhere'? Nonetheless, this does highlight the need for externalists to give a positive account of what normative reasons *are*, one that does not entail that knowledge of reasons requires us to transcend our subjectivity. I address this question in chapter eight.

Cowley also charges externalists of making the agent passive with respect to reasons for action. Cowley says that on the externalist account:

a normative reason for action... is a normative reason for action X to be *done* by A in circumstances C regardless of A's contingent psychological state or dispositions (2005 p. 359)

However, Cowley complains, this makes the agent "merely a causal conduit for the required performance of X" (*ibid.*). Supposedly externalism renders agents unacceptably alienated from their own actions, because all that matters is that particular actions get performed, actions with which the agent may not identify at all.

Again Cowley's complaints seem unfair. Externalism isn't the claim that all reasons must be impersonal and universal but the claim that reasons *qua* reasons are not dependent on the subjective motivations of agents. An externalist might quite consistently claim that everyone has a reason to pursue their own personal projects. Cowley seems to mistake the internalism-externalism debate for a debate about whether reasons can be based in personal particularities. Externalists can accept that there might be such reasons, whilst denying that their status as reasons is dependent upon being related through a sound deliberative route to the motivations of the agent. Cowley discusses the example of a man being cruel to his wife who really does not care and lacks any motivation that would be served by treating her better (2005 pp. 365-6 cf. Williams 1995 p. 39). He says that what we say about such an example can only be understood "against the background of the particular relationship between the man and his wife" (2005 p. 366). This is true, but it misses the point. The externalist claim is that we can say there is a reason for the husband to change his ways regardless of whether there is some rational deliberative route by which the man could become motivated to stop beating his wife. The reasons are based in how the wife deserves to be treated, not



in the man's motivations. This is what divides internalists and externalists, not the question of whether all of the details of the particularities of their relationship and the situation are relevant. Externalists are just as committed as internalists to thinking that these things matter.

Facts about what a husband would be like if he were more rational add nothing to our understanding of why he should treat his wife respectfully; it obscures the point by adding unclear<sup>26</sup> and irrelevant<sup>27</sup> information. The analysis mislocates the source of normativity<sup>28</sup>. According internalism, judgements like 'I have a reason not to beat my wife' are made true by facts like 'were I to be fully informed and rationally deliberate I would desire not to beat my wife'. This is wrong kind of fact to make true the judgement in question and to act as a source of normativity. *That I have a reason not to beat my wife* seems to be straightforwardly made true by facts about my wife (e.g. that she deserves to be treated with respect) and my relationship to her (e.g. that I have promised to love and cherish her). *These* facts make claims on my conduct. The internalist analysis obscures this fact rather than illuminates it.

Goldstein claims that "it is important that we appreciate the enormous strength of this assumption that an internal reason statement about this man is false" (2004 p. 73). Goldstein shows the strength of this assumption by mentioning various desires that could be fulfilled by refraining from beating his wife, such as the desire to stay out of jail. Since there are many such possible desires, it is possible that the husband possesses one of them. However, this misses the point. The externalist claim is not (merely) that all husbands have a reason to treat their wives with respect, but that the *source* of such reasons is not in the man's motivations but in e.g. what wives deserve. When we ask whether the man has any reason not to beat his wife, externalism allows us to answer immediately and unhesitatingly that her welfare straightforwardly provides a reason. Internalism, in contrast, has the perverse consequence that we must answer "it all

---

<sup>26</sup> It is unclear what is meant by 'more rational' until an analysis of rationality is given. It is also unclear whether it is possible to come up with a single determinate scale of rationality against which agents can be assessed as more or less rational.

<sup>27</sup> If we already know that the man's wife deserves to be treated respectfully, what is added by saying he would treat her respectfully if he were more rational? Even if the latter claim is true, it remains secondary to and dependent upon the original point – that he ought to treat her respectfully.

<sup>28</sup> In fact, it isn't clear whether this account succeeds in identifying a source of normativity at all. It isn't clear why we should care about satisfying the members of our actual or idealised S. Nor is it clear that we actually care about satisfying our S. As Audi says, although "any normal person wants many things, it cannot be seriously maintained that such a person thereby wants their conjunction or... has a second order want that all these first order wants be satisfied. I might, for one thing, be unable to get such a huge

depends on what he wants...” Normative reasons are supposed to be considerations that direct our actions. If they cannot even direct a man not to beat his wife without invoking some convoluted procedure invoking counterfactual motivations then something has gone seriously wrong.

## 7.8 Internalism and vice

In discussing the example of the cruel husband, neither Williams nor externalists such as Scanlon and McDowell take the man to be irrational. However, externalists can say things internalists cannot: that the man does not act as he should, that he ignores reasons that he has, and that he does what he has no reason to do. If internalism makes it impossible for us to justifiably criticise a man for beating his wife then clearly it is an implausible theory. Williams thinks that internalism allows sufficient resources for such criticism in the form of vice concepts. Williams writes:

There are many things I can say to this man: that he is ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal, and many other disadvantageous things (1995 p. 39).

However, as Scanlon (1998 p. 367) points out, the use of these vice concepts is only appropriate when particular kinds of *reason* are in play. If the husband was ungrateful then he had reasons to be more appreciative. If he was inconsiderate then he had reasons to take her feelings and preferences into account. However, according to internalists, there were no such reasons present, since the husband had no motivation that could render such reason statements true. Therefore, internalism entails that these attributions of vice to the husband are unjustified. Internalism makes it impossible for us to justifiably criticise this man for beating his wife. Therefore it is an implausible theory<sup>29</sup>.

Appropriately attributing virtue or vice to an agent depends quite generally on it being true that the agent succeeds or fails in responding appropriately to particular kinds of reason. Therefore, internalism suffers from a general problem of accounting for the justifiability of attributions of virtue and vice. The problem arises because in many cases vice is partially constituted by a lack of motivational responsiveness to reason-

---

conjunction before my mind and might in any event think it is unlikely to be consistent. And suppose I am an externalist. I may then think that I probably have some wants I should not have” (1996 p. 398).

<sup>29</sup> Cowley (2005 pp. 365-8) recognises this problem and makes a number of suggestions about the importance of context and relationships and about conceptual competence. However, nothing he says seems to be directly relevant to, or to contribute towards solving, this problem.

giving features. Internalism entails that in many cases of vice the lack of motivation entails the lack of a reason. The presence of a reason that the agent fails to respond to would justify our attributing the vice to them. However, on the internalist account the presence of the very feature of the agent that grounds the vice (i.e. whatever grounds the lack of motivation) prevents us from attributing a vice to them (by entailing there is no reason that they are failing to respond to). If there are no external reasons then cruel, inconsiderate and foolhardy people may have no reason to change, and their character traits may not involve an inappropriate failure to respond to reasons, since there may be no such reasons. Internalists might respond that rational deliberation must be conceived so widely that it includes all the processes that remove vice. However, now it is internalists who are guilty of ‘bluff’ and over-optimism about the capacity for rationality to guarantee reform towards moral ideals.

## 7.9 Internalism on epistemology and deliberation

Consider the example of a man, O’Brien, who enjoys throwing dinner parties but who always, halfway through dinner, starts telling inappropriate jokes that make his guests uncomfortable. His wife repeatedly explains to him that he should not tell such jokes because it makes his guests uncomfortable, but this makes no difference to O’Brien’s behaviour. At this point externalists might conclude that O’Brien has a reason not to tell inappropriate jokes: namely, that it makes his guests uncomfortable. However, internalism entails that we can only be justified in asserting this if we know (a) what O’Brien’s existing subjective motivations are and (b) what processes of rational deliberation he might engage and what their outcomes would be. This is problematic for two reasons. Firstly, it means our intuitions about when people have a reason might turn out to be badly mistaken. It might turn out that there is no process of rational deliberation and no subjective motivation that make it true that O’Brien has a reason not to tell inappropriate jokes. Yet we have a strong intuition that the fact that it makes his guests uncomfortable is just such a reason. Secondly, it makes it almost impossible for us to be justified in making assertions about reasons. Our reason judgements become tentative claims, awaiting empirical vindication that may never occur, or may occur years later. As Scanlon (1998 p. 365) says, there is an inherent difficulty in knowing what is grounded in an agent’s ‘S’. Moreover, Williams claims that there is “an essential indeterminacy in what can be counted a rational deliberative process” (1979 p.

25). If this is right it seems impossible for us to know what an agent would desire after rationally deliberating since we can never what their rationally deliberating would consist in<sup>30</sup>. The implausibility of the internalist epistemology therefore constitutes a powerful reason to reject internalism.

A further problem for internalism about reasons is in giving an account of rational deliberation. The internalist position presupposes that there are reasons for us to deliberate in some ways rather than others. This is obviously true, but it seems that internalists will have trouble accounting for it. To see why, we only need to ask, as Lillehammer does, whether reasons for deliberation are themselves internal or external:

Either what counts as a rationally sound deliberative route for an agent is relative to the content of his present desires, or it is not (2000 p. 510).

Lillehammer outlines three ways in which the criteria for rational deliberation could be dependent on the agent's desires. They might be deliberative processes that (1) the agent would desire to engage in if she were fully rational; (2) the agent desires to engage in; (3) would promote the agent's desires.

The first option is a non-starter: if we ask which deliberative processes are rational, it will not help to be told that they are the ones that rational people desire to engage in. Internalists then owe us an account of how it is that all rational people come to desire to engage in the same deliberative processes and of what these processes are. If rationality ensures conformance in desires about deliberation it can only be because there are rational norms that demand that we deliberate in some ways rather than others. And this returns us to the original question. The second option is even worse. It entails that any deliberative process can count as a rational. It also allows that anything that can be *believed* to be a deliberative process can count as a rational deliberative process. This is absurd.

The third option – the claim that rational deliberative processes are those that would promote the agent's desires – fares little better. According to this model, there are two ways in which an agent could have a reason to deliberate in a particular way,  $\Phi$ . The first is that the agent has a desire to deliberate in a particular way,  $\Psi$ , or deliberate to a particular conclusion,  $\chi$ , and believes that by deliberating in accordance with  $\Phi$  they

---

<sup>30</sup> Even if rational deliberation were determinate and we could know the contents of an agent's S, knowledge of reasons would probably *still* be almost impossible since it is likely that processes of rational

would deliberate in way  $\Psi$  or reach conclusion  $\chi$ . The second is that an agent may desire to perform a particular action,  $\alpha$ , or bring about a certain state of affairs,  $\beta$ , and believe that deliberating in accordance with  $\Phi$  would subserve  $\alpha$ -ing or bringing about  $\beta$ . However, in neither scenario does the agent's desire constitute a reason for them to deliberate in accordance with  $\Phi$  or make their deliberation in accordance with  $\Phi$  rational. The claim that it does depends on instrumentalism, and we have already seen that instrumentalism is false. Hence, this third internalist attempt to account for rational deliberation "entails a collapse into the sub-Humean model which the doctrine of internal reasons is designed to avoid" (Lillehammer 2000 p. 511)<sup>31</sup>.

There is no plausible account of reasons for deliberation that relates them to the desires of deliberating agents. Therefore, reasons for deliberation appear to be external. This casts doubt on internalism for two reasons. Firstly, deliberation is something that agents *do*. Even if deliberation is not a kind of action, it is similar to action and it seems true that deliberation requires reasons and motivations just as action does. Internalists owe us an account of why normative constraints on deliberation are independent of our desires, whilst normative constraints on action not only are not, but *cannot* be. Why should the idea of external constraints on action actually be *incoherent*, when, as internalists must agree, we can make perfect sense of external constraints on deliberation? If there is nothing incoherent about being able to respond to external reasons in general, why should they be incoherent in the case of action? If a normative requirement for me to deliberate in a particular way can motivate me independently of my subjective motivations, why should a normative constraint on action not be able to do so? It is difficult to see how there could be any answer to this question that is not based on the Humean theory of motivation. This makes Parfit's point all the more pressing when he writes:

Internalists must claim that, because some motivational fact obtains, something *else* is true: we have a reason for acting. In making that claim, they are committed to one kind of irreducibly normative truth. That undermines their reason to deny that there can be such truths about what is worth achieving, or preventing (1997 p. 129)

---

deliberation and agent's subjective motivations are highly complex and interact with each other in complex ways.

<sup>31</sup> The sub-Humean model is instrumentalism (Lillehammer 2000 p. 507). Williams says that "the sub-Humean model is certainly too simple" (1979 p. 18) and that his aim is to "work it up into something more adequate" (*ibid.*)

## 7.10 The conditional fallacy

Johnson argues that internalism commits the conditional fallacy, and that if it is reformulated to escape this fallacy it loses its distinctive appeal. The conditional fallacy “consists of overlooking, in various ways, dependencies between the *analysandum* and the antecedent and consequent of the conditional in the *analysans*” (Johnson 1999 p. 54). Internalists cannot account for the fact that we may have reasons to do things that are dependent on either the fact that we are not currently fully informed or the fact that we have not yet deliberated rationally. For example, it is not unusual for someone to have a reason to get better informed, but someone who is already fully informed could not have a reason to get better informed and would not desire to get better informed. Therefore, what we have reason to do cannot be dependent on what we would want if we were fully informed. Johnson writes:

Simply doing something to become more rational is an action of this kind. Were this of value to someone, [internalism] has the absurd implication that if the agent were not rationally lacking in any way, he would want to become more rational (*ibid.*).

Johnson illustrates this point with the following example:

[O]n some ill fated day you might come to believe that you are agent James Bond, licensed to kill, baccarat shark and connoisseur of fine port. And because of this belief you might come to have desires to play high-stakes baccarat, buy cases of rare wine and sneak around Russian embassies. Yet [internalism about reasons] implies that there would be no reason for you to seek a psychiatrist, since were you fully rational you would not have this false belief and so would have none of these irrational desires. (1999 p. 61)

Johnson points out that “few of us can say that our cognitive and imaginative powers could not be improved; and even the strongest willed among us can have difficulty staying on a diet” (*ibid.*). If we accept internalism about reasons we have to accept that there is “no reason for us to improve in these areas” (*ibid.*) since if we were fully rational we would not suffer from such deficiencies.

Some authors (Railton 1986; Smith 1995) reject internalism but embrace a modified view according to which “necessarily, a reason for A to  $\Phi$  entails that if A is completely rational, he desires, not himself in his perfectly rational state, but his less rational self, to  $\Phi$ ” (Johnson 2003 p. 574). The problem with this position is that it makes it difficult to see how reasons can explain or justify actions. Johnson writes:

Suppose, for instance, I become convinced that I am James Bond. Then although my more rational self desires that I should remove that belief, it is impossible for his desire to play any role in the explanation of my removing it. (2003 p. 575)

For me to come to believe that my more rational self would want me to not believe that I am James Bond I would first need to come to believe that I am not James Bond. Whenever I take myself to be James Bond and to have a reason to sneak around Russian embassies, my reflections about what my more rational self would want will start from these assumptions and yield the conclusion that I do indeed have a reason to sneak around Russian embassies. True, my more rational self would want me not to sneak around Russian embassies, but this fact can play no explanation in my coming to see that I have no reason to do so. Given my present cognitive limitations, I am unable to access truths about my rational self that are such that were I to believe them they would explain how I came to see what I really had reason to do (see a psychiatrist). In reality I have a reason to see a psychiatrist, but internalism entails that I have reason to sneak around Russian embassies.

Van Roojen suggests that internalism can avoid the conditional fallacy by making statements about the reasons that *A* has dependent on truths about the desires of a less than fully idealised version of *A*. Suppose we “let ‘rational<sub>R</sub>’ stand for ‘rational except in aspect *R* of rationality’” (Johnson 2003 p. 579) and include being fully informed as an aspect of rationality. Counterexamples to the original internalist position arose in cases when *A*’s reason to  $\Phi$  was dependent on *A*’s not being fully informed or fully rational in some way, *R*. Van Roojen’s proposal (2002 p. 238) is that in such cases we say that *A*’s having a reason to  $\Phi$  is dependent upon the fact that were they rational<sub>R</sub> they would desire to  $\Phi$ . However, as Johnson says, this fails to avoid the conditional fallacy. Any action that would make an agent more rational in every way would be an action they had a reason to perform. However, for such an action the rational<sub>R</sub> version of an agent would be identical to the actual agent. This entails, falsely, that agents have a reason to perform actions that make them more rational in every way if and only if they desire to perform those actions. If, on the other hand, we return to the original analysis in such a case, making the reason dependent on what a fully rational agent would desire, it still turns out that agents have no reason to pursue the action in question. Fully rational agents do not desire to become more rational. Either way, the internalist analysis fails.

### 7.11 Internalism doesn't satisfy the guidance condition

The appeal of internalism is that it is a way of making sense of some important ideas about the relationship between motivation and reasons for action. It is widely held that normative reasons must be capable of motivating and that awareness of normative reasons must necessarily motivate rational agents. That is, normative reasons must not only satisfy the following 'internalism requirement':

Practical-reason claims, if they are really to present us with reasons for action, must be capable of motivating rational persons. (Korsgaard 1996a p. 317)

They must also satisfy the following 'motivational requirement':

[I]f agent A has reason  $r$  to perform action  $x$ , and A is properly aware that  $r$  obtains, then A must be motivated to do  $x$ , on pain of irrationality (Wallace 1999 p. 218)

They must satisfy this latter requirement because rationality is, at least in part, a matter of responsiveness to reasons. Thirdly, normative reasons must meet the 'guidance condition':

[T]he motivations and actions of rational agents are guided by and responsive to their deliberative reflection about what they have reason to do (Wallace 1999 p. 219).

If this condition is not satisfied, we will "not be able to make sense of the idea that persons are genuine agents, capable of determining what they shall do through the process of deliberation" (*ibid.*).

The motivational requirement is often thought to provide an argument for internalism since it states that whenever there is a reason for someone to  $\Phi$ , and they are aware of this they must, if they are rational, be motivated to  $\Phi$ . It might seem natural to conclude that anyone who knows all the relevant empirical facts and deliberates rationally is rational. It then follows that if an agent who knows all the facts and deliberates rationally but is not motivated to  $\Phi$ , there is no reason to  $\Phi$ . However, accepting the motivation requirement does not force us to accept this conclusion. The argument might run the other way: from the premise that there is a reason that left the



agent unmotivated to the conclusion that the agent wasn't rational. That is to say, empirical knowledge and rational deliberation might be insufficient for rationality.

There is another way of thinking of the internalist argument, according to which it is based on the assumption that "[t]o be motivated in a certain way is to be subject to a desire of some kind, and desires in turn are psychological states with respect to which we are ultimately passive" (Wallace 1999 p. 222). Rejecting this assumption entails that internalism is false: if we can actively control our motivational states then we come to believe that we have reason to  $\Phi$  we can decide to adopt a motivation to  $\Phi$  independently of our prior motivations. If, on the contrary, we accept this assumption, internalism appears to be the only way of meeting the guidance condition:

The guidance condition can only be satisfied in the right way if it is possible for us, through the activity of reasoning, to acquire a motivation that corresponds to our views about what we ought to do. But motivation is not the kind of state that is under our direct control in this way. Hence it is not so much as possible for us to bring ourselves to be motivated appropriately solely through reflection on allegedly external reasons for action; practical reason can move us to action only if it taps into substantial motivating states to which we are already subject (*ibid.*).

However, the appearance that internalism meets the guidance condition is an illusion. If I am passive with respect to my desires then when I reflect on a reason that I have to  $\Phi$  and this reason motivates me to  $\Phi$ , I am passive with respect to the generation of this new motivation. Its generation is external to my agency; its generation occurs within me without my participation, it is not guided by my reflective choice. As Wallace says, on this account, "the motivational effects of reasons are not traceable to our deliberative activity as agents, but to the operation of causal forces within us" (1999 p. 223). Normative considerations can only apply in contexts where agents can control outcomes, where there is reflective choice. Therefore, if we are passive with respect to our motivational states then the guidance condition is not met. Any adequate theory of normative reasons must satisfy the guidance condition: normative reasons are not normative if they cannot guide us. Therefore, internalism is false.

## 7.12 Conclusion

We have seen that internalism and instrumentalism are unsatisfactory. This allows that there *might* be moral reasons that apply to us independently of our desires. However, we

have not yet done anything to show that there *are* any such reasons. Nor have we provided an externalist account of what normative reasons are that satisfies the criteria to which we held internalism accountable and overcomes the problems that internalism faced. It is to these tasks that I turn in chapter eight.

## Chapter 8. The Normativity of Morality

### 8.0 Introduction

I have shown that there *might* be moral reasons, but not that there *are*. Ordinarily it seems to us that there are moral reasons, but we can doubt whether moral considerations really are normative reasons. Such doubts may arise because of amoralists and cynics who ignore the demands of morality, or because of moral disagreement, or because morality is difficult, demanding sacrifices of us. In this chapter I aim to counteract such doubts by giving a Kantian account of moral normativity.

### 8.1 The normative question

When someone claims that a particular action is morally required we can legitimately ask why this is so: what is it about the action, its purpose and these circumstances that make it the case that morality requires this action? If I am morally required to attend a particular party there must be something that *grounds* this requirement: perhaps I have promised to go. However, the fact that justifying moral claims involves mentioning considerations that are called reasons does not show us that there really are moral reasons, that morality is normative. Our moral practice might be purely conventional, and deviating from it might not involve any kind of error. To put the point a different way, when particular considerations ground an obligation, we can ask *how* they do so. How is it that the fact that I have promised to go can provide me with a reason to go, whilst other facts (e.g. facts about what I had for breakfast) make no difference?

The problem under discussion goes by a number of different names: the problem of the amoralist (Raz 1997), the normative question (Korsgaard 1996b), the normative regress problem (Elstein 2008), the search for the foundations of morality (Williams 1985) or the Enlightenment project of justifying morality (MacIntyre 1987). The problem in each case is in establishing that moral demands are normative. In considering amoralists, for example, we are challenged to show that anyone who ignored moral considerations would be making some kind of mistake: they would be ignoring reasons that everyone has, reasons to take moral considerations into account. According to Korsgaard, this question becomes urgent “when what morality commands,

obliges, or recommends is *hard* [...] then the question – *why?* – will press, and rightly so. Why should I be moral?” (1996b p. 9)

This challenge cannot be met by saying that moral considerations are ordinarily taken to be reasons. The challenge is to show *what* makes moral considerations normative, *how* they are reasons, *how* they make justified claims on us. Dreier writes:

[S]uppose a moral theorist has proposed a certain set of rules, as the set of moral rules, the set that constitute morality... We could ask what reason we, or anyone, has to follow these rules. This would be a request for a practical justification (1997 pp. 81-2)

Alternatively, a justification might be requested for the claim that the proposed set of rules really is the set that constitutes morality. The prevalence of moral disagreement reminds us that this is of more than merely academic interest. The two challenges are linked: there is little point in showing that a particular set of rules is the set that constitutes morality if we have no reason to follow moral rules. Conversely, if we think that moral rules are normative, showing that a particular set of moral rules is normative also suffices to show that they are (at least part of) the set that constitutes morality. Providing a practical justification is the way to vindicate cognitivism and to deal with moral disagreement.

## 8.2 Realism about reasons

In asking the normative question we are asking whether morality gives us reasons, so it seems that what is required is a defence of realism about moral reasons. This returns us to our reflections on the nature of reasons of the previous chapter. There we rejected internalism about reasons. However, in so doing we only committed ourselves to a negative externalist claim: that the reasons for action that an agent has display no dependence on their prior subjective motivations. At least three externalist options remain open to us. Reasons for action might be beliefs, facts, or facts together with needs or interests. I reject the first of these views and suggest that the second and third views are both right as far as they go, but that more needs to be said if an adequate account of reasons for action is to be given.

Firstly, we might affirm that whilst reasons are not dependent on desires (and *a fortiori* not constituted by desires) they are constituted by other mental states of the agent. We might affirm psychologism about reasons. More specifically, we might think

that reasons are constituted by the beliefs of agents. Consider the following reason statement:

Since he believed that she was alone and in trouble, he should have offered help (Dancy 2000a p. 49)

We are tempted to say that this might be true (even) if his belief was false, and that therefore even when the corresponding state of affairs *does* obtain it is his belief, rather than the facts, that provides him with a reason to help. However, we can resist this temptation by proposing instead that he has a reason to refrain from simultaneously (a) believing that she is alone and in trouble and (b) not offering help. When he has a reason to help this is because she is alone and in trouble. When she is not alone and in trouble, this is a reason not to believe that she is alone and in trouble, and there is no reason to offer help. However, believing something that entails that there is a reason to offer help shows us why the action of the agent with the false belief is rationally intelligible, even though he does what he has no reason to do<sup>1</sup>. Consider two further examples that Dancy gives:

[T]hat I believe that the cliff is crumbling is my reason for avoiding climbing it, because having that belief I am more likely to fall off (I will get nervous) (2000a p. 124).

Someone who believes that there are pink rats living in his shoes may take that he believes this as a reason to go to the doctor or perhaps a psychoanalyst. This is quite different from the person who takes (his belief) that there are pink rats living in his shoes as a reason to call the pest control officer (2000a p.125).

The fact that there are pink rats living in my shoes *would* (if it were indeed a fact) give me a reason to call a pest control officer. Likewise, the *fact* that I believe there are pink rats living in my shoes (when, of course, there are not) provides a reason for me to see a psychiatrist, and this is true whether I *believe* that I have this belief or not. Hence, this reason also seems to be fact-based rather than belief-based. The fact that I have beliefs about pink rats may also give other people reasons to help me seek psychiatric help, independently of any beliefs that those people might have.

---

<sup>1</sup> Piller (2003) expresses some doubts about this account, but since I think that Dancy (2003b) deals adequately with these doubts, and because of limitations on space, I do not discuss them here. Nonetheless, it should be reiterated that even reasonable beliefs do not themselves give us reasons for action. Their reasonableness may be *indicative* of reasons for action that we have, and we may be *excused from blame* when our reasonable beliefs lead us to do what we should not do (cf. Piller 2003 p. 444).

We have moved from thinking of reasons as constituted by beliefs to thinking that they are constituted by facts. This is a tempting position, but it can, at most, be part of the story. We also need some account of what makes facts into reasons. There are many facts that rarely give us any reason for action. If I were married, this fact would give me reasons for doing a lot of things, but the fact that one of the moons of Jupiter is called Ganymede rarely provides me with any reason to do anything. Certain facts are significant, reason-giving. We need some explanation of what makes matters of fact into reasons. The appeal of internalism about reasons is that it provides such an explanation: internalists claim that matters of fact become reasons when they are suitably related to our prior subjective motivations.

Externalists might try to provide such an explanation, and account for the intuitions that make internalism appealing, by claiming that matters of facts become reasons when they relate to the needs and interests of human beings<sup>2</sup> (Lowe 2008). This in itself does not provide the account we were looking for, since an account is then required of why it is that needs and interests can ground reasons for action. They seem, after all, to be just facts of another kind: facts about human nature. A need is a necessity. Whenever something,  $\Phi$ , is needed, it is needed by something or someone, A, for something else,  $\Psi$ . For example, human beings (A) need food ( $\Phi$ ) for survival ( $\Psi$ ). Needs are constituted by three-place modal relations of this kind, and hence needs are modal facts. They are not only facts themselves, but they are also grounded in other facts – human needs are grounded in facts about human nature. For example, the human need to consume particular kinds of nutrients is grounded in facts about our metabolism. A need can be said to exist whenever any necessity of this kind exists, whenever any modal claim of the relevant kind is true. Needs vary in the kinds of reason for action that they provide; some provide no reasons for action at all. Therefore, some account of how and when needs provide reasons is required.

One way of accounting for this would be to say that a need provides a reason for us to  $\Phi$  to the extent that  $\Phi$ -ing is necessary for the realisation of some good. The problem with this proposal is that it leads immediately to a further set of questions about the nature of goodness about which we are unlikely to be able to give any satisfactory answer. As noted in chapter one, goodness seems to be linked to motivation and action, and it is natural to understand this by thinking of the features that ground goodness as

---

<sup>2</sup> The needs of animals and other beings may also play a similar role, but there is no space to discuss this here.

providing us with reasons to promote, protect, respect or realise certain things or states of affairs. If we understand reasons in terms of what is necessary to bring about goods then we will be unable to understand goodness in terms of reasons on pain of vicious circularity<sup>3</sup>. This will leave us with no way of making sense of the practical significance of goodness and its relation to action and motivation. We will still require some account of what it means for something to be good. Saying that goodness is a property is of no help; it is no better than saying that reasons are facts. We want to know what it is about some facts that makes them into reasons, and what it is about some things that makes them good.

To answer the question of whether there are moral reasons just by pointing to the facts that supposedly constitute those reasons is to affirm what Korsgaard calls 'substantive realism'. She writes, "the substantive moral realist thinks that there are correct procedures for answering moral questions *because* there are moral facts or truths" (1996b p. 36), that there are "intrinsically normative entities" (p. 33) whose existence cannot be questioned. Korsgaard criticises this position:

Having discovered that obligation cannot exist unless there are actions which it is necessary to do, the realist concludes that there are such actions, and that they are the very ones we have always thought necessary, the traditional moral duties. (1996b p. 34)

As Korsgaard says, "when the normative question is raised, these are the exact points that are in contention – whether there is really *anything* I must do, and if so whether it is *this*" (*ibid.*). Substantive realism is useless when we are faced with moral disagreement. If two apparently reasonable people or communities are engaged in a disagreement about whether some action,  $\Phi$ , is permissible it is of no help at all to merely insist that it is a *fact* that  $\Phi$  is (or is not) permissible. The same goes for attempting to answer the normative question by pointing to moral facts. Korsgaard writes:

It is because we are confident that obligation is real that we are prepared to believe in the existence of some sort of objective values. But for that very reason

---

<sup>3</sup> It might be possible to break out of this circle by giving an account of 'the good life' or 'the good for man'. As far as I can see, nothing I have said rules out this possibility. I suspect that if such an account can be given it would complement the account of morality that I give here according to which reasons are shared and we must never treat others as a mere means. It is likely that the good life is built on co-operation and mutual respect – the shared nature of reasons and the value of humanity would seem to be presuppositions underlying any claim that there is a good life for *us*, collectively, as human beings. However, there is not space to pursue this question here.

the appeal to the existence of objective values cannot be used to support our confidence (1996b p. 40).

Whether substantive realism is true or not, it cannot be used to answer the normative question. Substantive realism is just the claim that there are moral reasons, it is not an argument for that claim.

### 8.3 Normative regress

Externalism and realism do not, by themselves, provide an account of the normativity of morality. However, the situation is made even worse by an argument that threatens to show that no system of norms can possibly be justified. The argument in question is ‘the normative regress argument’. Elstein explains:

[S]uppose that some norm N is justified; we can ask what justifies it; but whatever the answer, it will presuppose that the fact cited justifies N, which is a further normative claim (since claims about justification are normative); so every justified norm depends on some further norm for its justification; thus there is a vicious regress of normative justification, and so no norms are justified. (2008 p. 4)

In attempting to block the regress we search for the unconditional, that which “makes it impossible, unnecessary, or incoherent to ask again” (Korsgaard 1996b p. 33). This suggests that only a transcendental account can show why there are any justified norms. Hence, intriguingly, the discovery of the regress that threatens to show that no norms are justified can become the basis for a Kantian argument that aims to justify moral norms. The argument can be stated by reference to a simple example. Suppose I decide to go to a particular party because I promised to. Saying that I went *because* I had promised indicates that I decided to go on that basis, and that in my deliberations about whether to go I took my promise to be something that favoured going (or even obligated me to go). I took that consideration to have normative force. Whenever we decide on the basis of a consideration we take that consideration to speak in favour, to be normative. However, in reflection we discover that we cannot find the unconditional that grounds the normative status of reasons. Nonetheless, we must still act, and our reflective nature means that when we act we must do so for what we take to be good reasons. This being the case, I must take the source of normativity to be in my reflective deliberative agency *itself* – for it cannot be found anywhere else, and it is my nature as a



reflective deliberating agent that demands that I find a source of normativity somewhere. Korsgaard writes:

[W]e still *do* make choices and have the attitude that what we choose is good in spite of our incapacity to find the unconditioned condition of the object's goodness in this (empirical) regress upon the conditions, it must be that we are supposing that rational choice itself *makes* its object good. (1996a p. 122)

Therefore, agency is the source of normativity – of reasons and obligations – and we have reasons and obligations to respect agency wherever we find it. Since you must see your power of rational choice as having a value conferring status, “you must view anyone who has the power of rational choice as having, in virtue of that power, a value-conferring status” (Korsgaard 1996a p. 123). In order to be capable of conferring value, “rational nature must itself be something of unconditional value – an end in itself” (*ibid.*), and therefore we must always treat the rational nature of any person as an end in itself. Therefore we have obligations to respect all persons, and these ground all our moral obligations, providing moral reasons and giving moral claims normative force.

I do not think that this argument is quite successful even if we accept that moral obligations can be derived from the claim that agency is the source of normativity. The argument encourages us to think that the normativity of the considerations that we take to be reasons is born out of a reflective crisis in which we postulate a source for normativity in reflective agency only because we cannot find it anywhere else. It also encourages us to think that each individual, having failed to find normativity in the world, independently and arbitrarily confers normativity onto the objects of his choice simply because he needs to see some things as normative and he cannot help valuing his own agency. I think we should not see things like this, since we can have (and respond to) reasons and obligations even if we never experience a reflective crisis in which we are driven to affirm agency as the source of normativity. It also seems unsatisfactory to see normativity as having its source in anything arbitrary; if rational choice confers value onto ends that *lack intrinsic worth* then normativity seems to have its source in *arbitrary* choices. These are not decisive criticisms of the argument just sketched but I think that we can construct a related argument based on the shared nature of reasons that gives us a more satisfactory account of the normativity of morality. In the next section I set out the structure of this argument. In that section and the one that follows I argue

that reasons are universal and shared in the ways the argument requires. I then return to the argument to explain what it shows.

## 8.4 Universal reasons and the categorical imperative

Korsgaard distinguishes between the categorical imperative and the moral law. The categorical imperative is “the law of acting only on maxims you can will to be laws” (1996b pp. 98-9). The moral law, on the other hand, “tells us to act only on maxims that all rational beings could agree to act on together in a workable co-operative system” (p. 99). Korsgaard aims to establish the normativity of the moral law by first establishing the normativity of the categorical imperative. Her argument (or, her reconstruction of Kant’s argument) for the categorical imperative is as follows:

1. A free will is “a rational causality which is effective without being determined by any alien cause” (Korsgaard 1996b p. 97)
2. “[B]ecause the free will is a causality, it must act according to some law or other” (*ibid.*; cf. Kant 1785 §446)
3. Given that the free will is not determined by an external (alien) cause, “the will must... have its *own* law or principle... until the will has a law or principle, there is nothing from which it can derive a reason” (Korsgaard 1996b p. 98)
4. The only constraint on the law that a free will makes is that it must have the form of a law, so the law of a free will is that “It must choose a maxim that it can regard as a law” (*ibid.*)

This argument is supposed to entail the categorical imperative: “I should never act except in such a way that I can also will that my maxim should become a universal law” (Kant 1785 §402). If the argument is sound, this imperative is binding on all rational beings, on all beings that can act freely on the basis of reasons.

To avoid questions about the relation between causality and law, and because I am adopting a practical reason approach, I focus on the notion of a reason for action. If we can show that when someone acts on a reason they act on a shared, universal consideration, we can use the relation between maxims and reasons to show that acting on a reason involves acting on a maxim that one can regard as a universal law. I argue

for this claim in three steps. Firstly, I claim that when I decide to act on the basis of a consideration I act on the basis of a reason that also applies on other occasions. Secondly, I claim that reasons for action are not private but are always shareable or shared. A consideration that is a reason for you can at the same time be a reason for me. Thirdly I claim that since maxims express relations between reasons and actions, one acts for a good reason only when one acts on a maxim that one can will as a universal law.

Firstly, then, I argue that deciding to act on the basis of a consideration involves acting on the basis of a reason that also applies on other occasions. From the practical point of view, when we are deciding how to act, we regard ourselves as free to determine our own actions on the basis of reasons. Acting for a reason is completely different from acting on a whim. I cannot take something to be a reason for me to  $\Phi$  now whilst at the same time taking it to be no reason to  $\Phi$  for someone else in exactly the same circumstances, or for me in relevantly similar circumstances. If I think that the fact that I promised is a reason for action for me now I cannot think that it is no reason at all in other circumstances unless there is some relevant difference – something that defeats the reason in those other circumstances. Otherwise I have no conception myself of what having a reason to  $\Phi$  consists of over and above my now  $\Phi$ -ing. I have no way of distinguishing between my  $\Phi$ -ing on a mere whim and my  $\Phi$ -ing on the basis of a reason. We must think of reasons as being universal if we are to distinguish them from whims.

We might now ask why we should not act on whims rather than on reasons. This is an interesting question. If I act on a whim then what motivates my action is not a normative reason for action – this point has just been established. Yet I am motivated by something – my whim. Is this a motivating reason or not? If so, I must, by definition, take this whim to be a good reason for action. But then I will be making a mistake and misrepresenting my own activity to myself. I will be trying to act on a reason and failing. This is something that, even by my own lights, I should not do. If, on the contrary, my whim does not constitute a motivating reason then it is difficult to see in what sense I am performing an autonomous action at all. If I do not think that my action, seen in the light of my whim, has anything to be said for it, then it is impossible for anyone, myself included, to make any sense of how it can be *my* action at all<sup>4</sup>. It is

---

<sup>4</sup> It might be that in some instances of having a whim I also have a reason, but this does not mean that I have a reason *in virtue of* having a whim.

impossible to see how the event in question is a product of my reflective deliberation and agency.

It might be thought that I can act on a whim or a desire, or even on a reason “without committing myself to acting in the same way whenever I have this desire” (Korsgaard 1996b p. 227) (or this whim or perceived reason). However, this line of thought overlooks an important fact:

[W]illing is *self-conscious* causality, causality that operates in the light of reflection. To will is not just to be a cause, or even to allow an impulse in me to operate as a cause, but ... [to] make *myself* the cause of what I do. And if I am to constitute *myself* as the cause of an action, then I must be able to distinguish between *my* causing the action and some desire or impulse that is ‘in me’ causing my body to act. (Korsgaard 1996b pp. 227-8)

For something to be identifiable, even to myself, as an act of willing, as something that *I* do, it must be based on reasons, on considerations that *I* take to favour action in a regular and non-arbitrary way.

The non-arbitrary nature of acting for a reason is connected to the phenomenon of supervenience. When two things have the same natural properties they must also have the same moral properties (see §1.6). Likewise, when two things have the same natural properties, they must have the same normative properties more generally – or we might say, when two things have the same natural properties they provide the same reasons for action. This means that nothing can be a reason for action whilst necessarily only applying on a single occasion. A reason for action must be a reason on every occasion when the features that ground the reason are present (assuming the same defeaters and enablers are also present). When I see myself as having a reason *because* (e.g.) I promised, I see promising as reason-giving not just here but more generally. I see that there is something about promising that provides reasons<sup>5</sup>. Of course, I may change my mind about reasons. However, just as I cannot act whimsically, I cannot change my mind about the reasons I have just on a whim. Korsgaard writes:

If I change my mind and my will every time I have a new impulse, then I don’t really have an active mind or will at all – I am just a kind of location where these impulses are at play... [T]his is not to say that I cannot ever change my

---

<sup>5</sup> Given the absence of certain defeating conditions and the presence of certain enabling features. E.g. if one only promised because one was deceived or manipulated then the fact that one promised might not provide any reason for action.

mind, but only to say that I must do it for a reason, and not at random. (1996b p. 232)

If I change my mind whimsically or at random about the reasons for action that I have then there is no sense in which it is really *I* that changes my mind. When *I* change my mind I must see the change in a positive light, I must make the change on the basis of some considerations that I take to favour the change. We can say, therefore, that if my actions are to be mine then they must be based on reasons rather than mere whims or desires, and that I cannot change my mind about the reasons I have on a mere whim or desire. Reasons are universal in the sense that I have the same reasons for action whenever the same features occur that ground that reason. In the next section I argue that reasons are also universal in the further sense that they are sharable by all agents. Reasons are not subjective, applying only to single individuals to whom they belong.

## 8.5 Shared reasons

In chapter seven I mentioned Smith's argument against internalism about reasons. Smith claims that we know that our concept of a reason is not relative because if it were then "we should expect to find that we are sometimes able to dissolve apparent disagreements by finding that both parties are speaking truly" (1994 p. 168). But disagreements about reasons for action are never resolved in this way.

If reasons are not relative then they are shared. When two people disagree about whether something is a reason to  $\Phi$  they are disagreeing about the presence of a single shared reason, not having a merely apparent disagreement about two separate things that they are conflating – their own personal reasons. If reasons are relative then a consideration, *C*, might constitute a reason for *A* to  $\Phi$  (reason<sub>A</sub>) whilst not constituting a reason for *B* to  $\Phi$  (reason<sub>B</sub>). In conflating the question of whether *C* constitutes a reason<sub>A</sub> and/or a reason<sub>B</sub> with whether it constitutes a reason to  $\Phi$  *per se*, *A* and *B* mistakenly think they are disagreeing. Their mistake is based on the false assumption that something could constitute a reason to  $\Phi$  *per se*. Once they understand that reasons are relative they will realise that such disagreement could never be possible. They might also come to see that they are both right: that there is a reason<sub>A</sub> to  $\Phi$  but no reason<sub>B</sub> to  $\Phi$  (or vice versa). Since we know that this is not what happens, we know that our concept of a reason is not the concept of something agent relative but of something shared. In the absence of extremely compelling arguments that support a universal error theory

about our reason-talk, we can justifiably conclude that reasons for action must be shared.

Other considerations also support the view that reasons for action are shared. The fact that we condemn some forms of talk as manipulative or deceptive presupposes that other forms of talk are not so. If I try to manipulate someone then I try to get them to act on the basis of something that I do not myself consider a reason<sup>6</sup>. If I am trying to persuade someone to  $\Phi$  without manipulating them then any consideration I offer to them must be one that I myself take to be a reason to  $\Phi$ . My aim must be to try to get them to also see this consideration as a reason to  $\Phi$  and to therefore give this weight in their deliberations about what to do. Therefore, if we see some forms of talk as manipulative, we are committed to the view that other forms of talk, not being manipulative, involve exchange of genuine reasons. Reasons can only be exchanged if they are shared. Therefore, the possibility of manipulation presupposes that reasons are shared.

We know that reasons are shared because we are always exchanging them. As Korsgaard points out, “[i]f ... reasons really were essentially private, it would be impossible to exchange or share them” (1996b p. 135). Korsgaard gives the following example of an exchange of reasons:

A student comes to your office door and says: ‘I need to talk to you. Are you free now?’ and you say ‘No, I’ve got to finish this letter right now, and then I’ve got to go home. Could you possibly come around tomorrow, say about three?’ And your student says ‘Yes, that will be fine. I’ll see you tomorrow at three then.’ (1996b p. 141)

Korsgaard offers two interpretations of what is happening. Firstly, if reasons are shared, they are “reasoning together, to arrive at a decision, a single shared decision, about what to do” (*ibid.*). If reasons are private then each of them “backs into the privacy of his practical consciousness, reviews his own reasons, comes up with a decision, and then re-emerges to announce the result... [T]he process stops when the results happen to coincide” (*ibid.*). The latter interpretation of events involves seeing others not as sources of reasons, but merely as features of the circumstances to be controlled or worked around. This is not how we generally relate to others; it is how we relate to others when we do not treat them as persons but instead try to objectify them and make them

---

<sup>6</sup> Similarly, I may try to get them to act on the basis of a consideration that I do think is a reason, but without being concerned to get them to autonomously act with this *as* their reason.

subordinate to our own plans and projects. Arguably a student must treat their teacher as someone with whom they share reasons, since learning from someone involves taking them as a source of reasons – reasons for belief and practice. More generally, it appears to be inconceivable for anyone to think of reasons as essentially private. I cannot think of myself as only<sup>7</sup> having private reasons – reasons that are not reasons for anyone else – because doing so involves willing that others not take my reasons into account. This in turn involves willing that they treat me as a mere means. This is incoherent, for to will that I be treated as a mere means is to will that my will not be effective, to will that my will be ignored. This is to will that others do something that they cannot possibly do – to take into account my wish to never have any of my wishes taken into account.

It might be thought that thinking of reasons as shared fails to make sense of a very important aspect of our experience and activity: our personal projects and ambitions. It seems that if I have an ambition – perhaps I am determined to climb Kilimanjaro – this provides me with reasons to pursue this ambition, but it does not give anyone else a reason to help me. On this view, it must be that there is no objective value in climbing Kilimanjaro, and in having my ambition I do not take this activity to have any objective value. If climbing Kilimanjaro had objective value then others would have reasons to help me which, *ex hypothesis*, they don't. If my reason for climbing Kilimanjaro was that it has objective value then I would be mistaken, and I would have no reason to climb but, *ex hypothesis*, I do have a reason – a reason that is provided by my ambition and not shared by anyone else. If we accept this account of ambitions then we might have to accept that reasons do not have to be shared. However, I do not think we have to accept this account of ambitions. People do not generally adopt projects that they take to be worthless independently of their adopting them. There are many things, for example, that can be said in favour of climbing Kilimanjaro. Korsgaard writes:

[A]sk a mountain climber why she climbs and she need not be mute: she may tell you things about the enlarged vistas, the struggle with the elements, the challenge of overcoming fears of surpassing physical limitations. She takes her desire to climb mountains to be a motivated desire, motivated by recognizably good features of the experience of climbing. She does not take the value of the climb to be conferred on it simply by her desire to do it (1996a p. 289).

---

<sup>7</sup> Perhaps I have not ruled out that some reasons might be private. However, reasons are not just considerations on the basis of which we decide. We also take them to justify our choices. Justification seems to be essentially interpersonal: I justify myself to others. We do seek to justify our choices to ourselves, but when we do so our justifications do not differ in form to those we offer to others: they involve giving the same kind of reasons. Since such reasons are those that can be exchanged with others, they are shared reasons. This is why I take all reasons to be shared.

Personal projects and ambitions generally display these features. As Korsgaard says, “to have a personal project or ambition is not to desire a special object which you think is good for you subjectively, but rather to want to stand in a special relationship to something you think is objectively good” (1996a p. 288). We might draw a comparison with what Dancy says about inclinations (see § 7.3). When we have personal ambitions and projects we have perfectly good shared reasons for what we do, it is just that these reasons are not conclusive. They do not generally obligate us to engage in these projects or make it the case that everyone must adopt them – for one thing, there are far too many valuable projects and pursuits for a person to adopt in a single lifetime. We are forced to choose to pursue only a few of the vast number of worthy projects available to us, and often we may have to choose between projects that are equally worthy or appear equally worthy given our limited perspective on them. That personal preference must play a role does not mean that we lack good reasons for what we do. We have good reasons, but they are not decisive. If you have decided to climb Kilimanjaro then it may be that I *do* have reasons to help you – and reasons to go along myself<sup>8</sup>. Usually, however, such reasons will be outweighed or defeated by the fact that I am, or could be, involved in so many other worthy projects, and perhaps ones that better suit my circumstances and personal strengths.

## 8.6 Maxims and the moral law

If reasons are universal and shared then any reason that I act upon is a consideration that could be a reason for myself or for anyone else on this or another occasion. We can think of a maxim as relating a reason to an action. Korsgaard says, “a maxim has two parts: the act and the end” (1996b p. 108). Suppose I jump into the water in order to save a drowning child. My action, jumping, is related to my reason, that the child is drowning, by something for the sake of which I act – in this case, saving the child. When this act is performed for this reason the maxim on which I act is ‘I will jump into the water to save the child’. When I act for a reason, I must hold that my reason would also be a reason for me or for anyone else in relevantly similar circumstances. In willing the action for this reason, I also normatively commit myself to relevantly similar actions

---

<sup>8</sup> Arguably there are many things that we have much better reasons to do. Instead of climbing Kilimanjaro we could be working to make the world more just and to alleviate poverty and suffering. Nonetheless, for



at other times and by other persons. In action I affirm that  $r$  is a reason to  $\Phi$  for me in circumstances  $C$ , and in so doing I affirm that  $r$  is a reason to  $\Phi$  for anyone in  $C$  (where  $C$ ,  $r$  and  $\Phi$  refer to *types* of relevantly similar circumstances, considerations and action). Korsgaard discusses Plato's example of the three maxims:

1. I will keep my weapon, because I want it for myself.
2. I will refuse to return your weapon, because I want it for myself.
3. I will refuse to return your weapon, because you have gone mad and may hurt someone (1996b p. 108).

Since each of the good maxims (one and three) has either the same action or the same purpose as the bad maxim (two) the goodness of a maxim is not in its parts. Instead, "if the action and purpose are related *so that* the maxim can be willed as a law, then the maxim is good" (*ibid.*). The idea is that we can will that everyone keep their own weapons whenever they want them for themselves. We take owning something and wanting to retain it to be a reason for retaining it, and in reflection we see that this can function as a universal shared reason. The same goes for not returning weapons to those who have gone mad<sup>9</sup>. However, when I want to keep something that is yours I cannot take this consideration to be a good reason for action. Taking it to be a good reason for action involves seeing it as something that would favour anyone acting similarly in similar circumstances. However, I cannot see your desire to retain my property as being a good reason for you to retain it. If I see something as mine then necessarily I must will that someone else merely wanting to retain it is insufficient to ground any claim to retaining it. If I do not then there is no sense in which I think of the item as *mine* since I see others as having equal claim to it. Since I cannot will that merely wanting to retain something that belongs to someone else can function as a universal shared reason to retain it, I cannot act on this as a reason at all. In refusing to return another's property merely because I want to keep it, I allow this mere desire to control my activity rather than actively taking control myself and acting on the basis of something that I can affirm as a valid reason for action.

---

the sake of argument I ignore here the fact that pursuing personal projects of this kind might be a luxury that, given the current state of the world, cannot justifiably be indulged in.

<sup>9</sup> Therefore, I should accept that if I own a weapon and it is mine to retain, then were I to go mad, others would be justified in taking the weapon from me. This seems right.

I have argued that the nature of acting for a reason allows us to see why we must conform to the categorical imperative. However, a distinction was drawn earlier between the categorical imperative and the moral law. The categorical imperative tells us to act only on maxims that we can will as universal laws. The moral law tells us that we must always treat persons as ends in themselves and never as a mere means. I now argue that we can use the categorical imperative to show why we must accept the moral law. If I must act on the basis of a universal shared consideration, then any action that I perform must be one that you can consent to since it must be one based on a reason that you can endorse. I must act on something that could also be a reason for you, something that either of us can will as a universal law for governing actions in the circumstances in which I now find myself.

Korsgaard argues that if we are to see ourselves as having “any reason to do one thing rather than another ... as having any reason to live and act at all” (1996b p. 121) we must treat our nature as reflective, deliberating agents as grounding our practical identities. We must endorse and acknowledge our rational deliberative nature in and through our action. Korsgaard suggests that this involves valuing ourselves as human beings. If I see my reflective deliberative nature as worthless, as something that I do not endorse, then I can no longer see myself as having any reason to do anything since reflecting on what I have reason to do would involve endorsing my reflective nature. I must value my own reflective, deliberative nature. This entails that I have a moral identity and moral obligations. Or rather, as Korsgaard says, “it does if valuing humanity in your own person rationally requires valuing it in the persons of others” (1996b p. 121).

However, is it the case that valuing ourselves commits us to valuing others? This may sound unconvincing. The sceptic might imagine that we are being urged to value others if we value ourselves because otherwise we will be *inconsistent*. If this is so, then some reason needs to be given to show that we really must be consistent in our evaluations. Even though we might be able to answer the sceptic on this point, this is not exactly how the argument should be understood. Since reasons are shared, and not private, any reason that I have to value humanity in myself is a reason that is shared by everyone. The argument is that *we* cannot act without valuing *our* identity as human beings. Our rational nature is shared, and it generates shared reasons that ground moral requirements, based on the practical necessity of valuing humanity. Korsgaard writes:

Suppose that we are strangers and that you are tormenting me, and suppose that I call upon you to *stop*. I say: ‘How would you like it if someone did that to you?’ (1996b p. 142)

I recognise that were you to torture me I would resent it and think that you have not just a reason, but an obligation, to desist. I would see this obligation as grounded in my objection to your action, in the fact that I do not consent to it and that you proceed without any concern for my wishes or well being.

Korsgaard expresses this thought by saying that I make myself an *end* or *law* for others. She says, “if you are a law to others in so far as you are just human, just *someone*, then the humanity of others is also a law to you” (1996b p. 143). I make myself a law to others just insofar as I am someone, since I claim that the other has an obligation to desist from torturing me even if I am a complete stranger. However, in making myself a law for others I give reasons and obligations directly to others. The reasons why you should stop torturing me are e.g. that I object to it, that you are causing me pain, that you are harming me. It is not that you have a subjective reason to be consistent and that you can only be consistent by refraining from torturing me. That would make me a mere means to your (private) goal of achieving consistency. You are obligated to treat me as an end, as a direct source of (shared) reasons. It is not that I first discover that I have subjective reasons to value myself and then that I have reasons to value you only because it would be inconsistent of me not to do so. Rather, since reasons are shared, the nature of humanity gives all of us reasons to treat persons as ends in themselves. It is the fact that we are able to object to how others treat us, and to make up our own minds about the role we will play in actions that involve us, that supplies all of us with obligations. Korsgaard writes:

Suppose you could say ‘someone doing that to *me*, why that would be terrible! But then I am *me*, after all.’ Then the argument would fail of its effect, it would not find a foothold in you. But the argument never really fails in *that* way. (1996b p. 143)

As Korsgaard says, “[i]f I listen to the argument at all, I have already admitted that each of us is *someone*” (1996b p. 143). In admitting this I admit that you can give me reasons, that we share reasons. Arguments, when they are good ones, give us reasons. If someone can offer me an argument then they can give me reasons.

The moral law is the law that we must never treat another as a mere means. Korsgaard explains this idea as follows:

According to Kant, you treat someone as a mere means whenever you treat him in a way to which he could not possibly consent... If I am forced, I have no chance to consent. If I am deceived, I don't know what I am consenting to. If I am coerced, my consent itself is forced by means I would reject<sup>10</sup>. (1996a p. 295 c.f. Kant 1785 § 430)

In contrast, treating someone as an end involves respecting “his right to use his own reason to determine whether and how he will contribute to what happens” (1996a p. 296). It is the shared nature of reasons that makes the moral law binding on all of us. Since we are all members of one community of reflective beings that can give and ask for reasons, our interactions must be governed by the moral law.

## 8.7 Second nature

The idea that reasons are shared, and the contrast between rational causality and the causality of nature, links to a theme found in the work of McDowell that can be used to strengthen the Kantian position. McDowell (1994) contrasts the ‘space of reasons’ with ‘disenchanted nature’, mirroring the Kantian distinction between ‘laws of nature’ and ‘laws of freedom’ (Kant 1785 § 387). If we adopt a theoretical standpoint in which we are interested in characterising the world objectively, as it is anyway, independent of all human concerns, then we develop a ‘disenchanted’ view of nature in which we concern ourselves with causal regularities rather than rational regularities, norms and reasons. From this point of view we see human beings as continuous with the rest of nature, as animals whose nature and behaviour we attempt to explain and predict according to causal regularities. This is not the perspective that we typically adopt in everyday life. Fully and continuously occupying this perspective is completely incompatible with having human relationships, with pursuing our day to day activities and projects and with the deliberative perspective in which we reflectively determine our actions. In action we see ourselves under the laws of freedom, occupying the space of reasons.

---

<sup>10</sup> Geoffrey Scarre suggested to me that this does not quite capture the point because someone of slavish temperament might be disposed to do whatever you ask of them even when you make requests that you shouldn't make, requests that exploit or degrade him. Even when a person of slavish temperament consents they might still end up being used as a mere means. Korsgaard's way of explaining the point remains illuminating, but we should remember that securing consent is insufficient to ensure that one does not treat someone as a mere means.

Once we distinguish between the ‘space of reasons’ and ‘disenchanted nature’ we come upon a problem when we try to find how these two perspectives fit together. We find ourselves asking how *values* can have any place in a world of *fact* and how human beings can be *rational animals*: how we can be a part of nature and yet have values, which seem not to have any place in the natural world. The temptation is to suggest that human beings have a dual nature, having a part that resides in the natural world and a part that transcends the natural and enters into a world of values. This is the temptation to ‘rampant Platonism’ in which “the structure of the space of reasons... is simply extra-natural” (McDowell 1994 p. 88) and our engagement with reasons and values is mysterious, “as if we had a foothold outside the animal kingdom, in a splendidly non-human realm of identity” (*ibid.*). We can resist the temptation to give such an account by seeing our rational nature as second nature, as something that “actualizes some of the potentialities we are born with” without introducing “a non-animal ingredient into our constitution” (*ibid.*).

Our second nature arises through processes of socialisation and formation that constitute “an element in the normal coming to maturity of the kind of animals we are” (*ibid.*) and that initiate us into structured practices that already exist prior to our individual contact with them. Hence, “[m]eaning is not a mysterious gift from outside nature” (*ibid.*). It is because I was spoken to as a child that I have a native language, because I was taught manners that I know how to be polite or rude, and know when I have reason to be offended by the rudeness of others. It is through the development of second nature that we come to have the conceptual capacities to reflect on our actions in the light of reasons, and to see ourselves as choosing on the basis of considerations that obligate us or favour our acting in particular ways.

Since we must undergo a process of formation in order to be able to develop the second nature that allows us to exchange reasons and through which “meaning comes into view” (McDowell 1994 p. 88), we initially have little authority over our engagements in the space of reasons. As a child I learn what is a reason for what from others who already know. What this shows us is that the development of our rational nature relies upon and already presupposes our upbringing in a community of persons, beings who exchange reasons, who engage in meaningful projects and set ends for themselves. I cannot develop as a rational being without accepting the moral law. The moral law is not only a precondition for the existence of the society that imparts second nature to me, but is also presupposed by my ongoing acceptance of the authority of

others to give me reasons and induct me into a meaningful social world. I only become an end in myself – that is, I only become capable of willing as opposed to following impulse and inclination – because others facilitate the formation of my second nature as a reflective, rational being. I cannot therefore deny their humanity, their nature as reflective, reason-responsive, purposive beings, for it is only through them that I was able to develop this nature myself. It is for this reason that McDowell says that the rational demands of ethics require no *external* justification, yet they are “essentially within reach of human beings” (1994 p. 84). Moreover, it is ethical upbringing that allows human beings to be initiated into the space of reasons, instilling “the appropriate shape into their lives” (*ibid.*), producing “habits of thought and action [that] are second nature” (*ibid.*).

Once we have been initiated into the space of reasons, however, we can come to ask the normative question, to wonder how it is that the claims of morality are justified. Our reflection on the nature of reasons themselves can in part answer this question: by seeing that the space of reasons is something that we necessarily occupy together, we see that our actions cannot be based on considerations that cannot be accepted as reasons by those that they affect. Korsgaard writes:

‘How would you like it if someone did that to you?’ In asking me this question my victim demands that I either cease using him as a means, or give up my claim not to be so used by others. But the latter is impossible: one cannot consent to be used as a [mere<sup>11</sup>] means. And so *he obligates me* to desist, and to treat him instead as an end in himself. (1996a p. 299)

## 8.8 Reasons and obligations on the practical reason approach

In chapter one I set out the practical reason approach according to which I aimed to show that morality is practical and moral knowledge is possible. Throughout I have been thinking of reasons as contributory, such that it can be the case that I have a reason to perform one action whilst having most reason overall to perform another. I have been assuming that moral reasons conform to this pattern. Now that I have given an account of the normativity of moral considerations, I need to consider how it links up with this practical reason approach.

---

<sup>11</sup> Korsgaard writes “as a means”, but her view is not that one cannot consent to be used as a means, but that one cannot consent to be used as a mere means. I consent to be used as a means whenever I intentionally help someone. If you are using me as a mere means then I cannot consent because I have no say in what is happening; I can attempt to consent but it will make no difference to you or to the nature of the action; I am deluding myself if I think otherwise.

The account of normativity we have given seems to yield more than one kind of reason. Firstly, there are considerations that I have simply learnt to take as reasons through the formation of my second nature. Some of these are such that were I to reflect on them I might be unable to give any account of why they constitute reasons. Some are such that after reflection I would no longer see them as reasons. There are other considerations that may show up to me as significant because of desires I have or ends that I have adopted. I may reflectively choose to endorse these considerations as reasons. The status of such considerations as reasons can only withstand reflection if they are able to function as a universally acceptable basis for action in the way required by the categorical imperative. A third kind of reason is generated when a maxim fails the categorical imperative test because it cannot be willed as a universal law. Reasons for keeping promises can be generated in this way: since promise-breaking maxims generally fail the categorical imperative test we generally have reasons to keep our promises. Such reasons are of a different kind to those of second nature and those brought about by reflective endorsement. Unlike these other kinds of reasons they generate *obligations*. As Korsgaard says, “[i]f reasons arise from reflective endorsement, then obligation arises from reflective *rejection*” (1996b p. 102). Moral reasons are reasons that arise in this way, since moral reasons take the form of obligations<sup>12</sup>.

There is *prima facie* something awkward about the account just offered. On this account it seems that moral reasons are *contributory* and that they constitute *obligations*. Even if it seems plausible that there can be contributory obligations such that we can be to some extent obligated to  $\Phi$  whilst overall being most obligated  $\Psi$ , this looks quite problematic in the light of the Kantian account of how obligations are generated. An obligation to avoid  $\Phi$ -ing on the basis of some consideration,  $r$ , or as a way of  $\Psi$ -ing, is generated when the maxim ‘I will  $\Phi$  because of  $r$ ’ (or ‘I will  $\Phi$  as a way of  $\Psi$ -ing’) fails the categorical imperative test. I have an obligation to repay money that I have borrowed and promised to repay even when I am in need because the maxim of not repaying promised money when I am in need fails the categorical imperative test:

[T]he universality of a law which says that anyone believing himself to be in difficulty could promise whatever he pleases with the intention of not keeping it would make promising itself and the end to be attained thereby quite impossible,

---

<sup>12</sup> This is not to say that obligations exhaust the category of the ethical, since we can consider ethics to be more general than morality, encompassing all considerations relating to what it is to live well.

inasmuch as no one would believe what was promised him but would merely laugh at all such utterances as being vain pretences (Kant 1785 § 422).

Treating another as a mere means to relieve one's financial difficulties is impermissible. In making an insincere promise one necessarily intends to make it impossible for one's benefactor to consent to one's action, for one aims to conceal from one's benefactor the true nature of the action. One cannot consent if one does not know what one is consenting to.

Since the obligation not to make insincere promises in order to get out of difficulty is grounded in this way, it is difficult to see how it can be only a contributory reason. However, we can overcome this difficulty, and do justice to the contributory, by seeing contributory reasons as *constituents* of maxims. Suppose I am considering borrowing money to get me out of financial difficulty. The details of my financial difficulties – e.g. that I am behind in paying my rent, that I need to buy food, that I have reached the limit of my overdraft – are reasons (and enablers for those reasons). The fact that you are willing to lend me the money that I need and the fact that I do not expect to be able to repay it are also reasons (for and against borrowing the money respectively). *Together* they constitute the contents of my maxim: given the facts of my circumstances  $C$  ( $f_1, f_2, \dots, f_n$ ) I will  $\Phi$  because of  $\Psi$ . This yields an overall verdict about whether or not, in  $C$ , I may  $\Phi$  because of  $\Psi$ , and hence whether I am obligated to not- $\Phi$ . Moral reasons are contributory, whilst the categorical imperative test yields overall moral verdicts<sup>13</sup>. A reason is a moral reason to  $\Phi$  or not- $\Phi$  in circumstances  $C$  whenever it is one of the conditions that contributes to the verdict delivered by the categorical imperative test that  $\Phi$ -ing or not- $\Phi$ -ing is impermissible.

---

<sup>13</sup> Dancy worries that using the categorical imperative to test whether we may act on a maxim may be incompatible with moral reasons being contributory (2004 pp. 68-70). We need to be able to make sense of the idea that moral reasons can be outweighed whilst retaining their normative force: that we can have good reasons to do one thing, but most reason overall to do something else. What I have said here is a way of dealing with this worry. Piller (2006) also addresses this issue, beginning from Korsgaard's (2008 pp. 207-231) Kantian insight that the nature of an action cannot be determined independently of the reason for which it is performed. Piller says that although Korsgaard "does not talk about a conflict of reasons... [w]e can, however, easily extend her account to appease Dancy's legitimate worry" (Piller 2006 p. 90). My action is not merely my  $\Phi$ -ing, but my  $\Phi$ -ing on the basis of  $R$  – this is why its moral permissibility is determined by assessing the relation between  $\Phi$ -ing and  $R$ . Likewise, if I  $\Phi$  on the basis of  $R$  despite  $Q$ , the relation between  $\Phi$ ,  $R$  and  $Q$  can be considered to properly describe the action. We can therefore see  $Q$  as a contributory reason and playing an important role in the normative picture. By reflecting on situations in which  $Q$  would, with the help of other considerations, have outweighed  $R$ , making it the case that one ought not to  $\Phi$ , we can understand the normativity of  $Q$ . My position is similar to Piller's except that I do not specify that we should insist that  $Q$  is part of the nature of  $\Phi$  in such cases.



To see that moral reasons remain contributory under the proposed Kantian theory, consider the role played by the consideration ‘that I promised to  $\Phi$ ’. Suppose there are no other relevant considerations, and I wish to break my promise but I am wondering whether it would be morally permissible. My maxim is ‘I will break my promise because I wish to’. This clearly fails the categorical imperative test (for the reasons Kant gives, above). Suppose, however, that I have promised to cut my mother’s lawn this morning but my friend urgently needs to be driven to the hospital. Now there is no difficulty in universalising a maxim that involves breaking a promise. Or suppose that I have accidentally made two incompatible promises. I have promised to meet a friend in town but I have also promised to wait at home for a delivery. Now I must weigh up the considerations on either side, see how I can rearrange my activities, see whether I can be released from one of my promises, and decide on the basis of all these contributory considerations taken together. If I am morally competent and experienced then I may find this easy and find it unnecessary to reflect much upon the source of my obligations and their relative strengths in order to get things right. However, in cases where I am unsure, I can reflect upon whether the maxim of my proposed action is a permissible one. This may be difficult in complex cases – but the moral life is often beset with these kinds of difficulties and our account of morality ought to reflect this fact. Even if the normativity of moral considerations comes from the categorical imperative and the moral law, experience and practical wisdom can often be more telling than explicit reference to the source of our moral obligations.

## 8.9 Internalism vindicated

In chapter two I claimed that if it could be shown that there are moral reasons then a moral internalism based on the practical reasons approach would be vindicated (§ 2.6). Let us see why. If morality gives us reasons then whenever there is a moral reason for me to  $\Phi$  there is a reason for me to  $\Phi$ . To believe that a particular consideration is a reason for action is to believe that I should take it into account in deliberation about what to do. If one sincerely believes that one should take a consideration into account in deliberating about whether to  $\Phi$  then either one *does* take that consideration into account or one does what one believes one should not (from the point of view of reason) do, believing oneself unreasonable in what one does. One is irrational, even by one’s own lights. Since morality does give us reasons, when I understand morality, I

understand that it gives us reasons. Therefore, if I understand morality, and I judge that  $r$  is a moral reason to  $\Phi$ , I believe that  $r$  is a consideration that I should take into account in deciding whether to  $\Phi$ . Therefore, when I understand morality, and judge that  $r$  is a moral reason to  $\Phi$ , I *do* take  $r$  into account in deliberating about whether to  $\Phi$ , unless I am irrational by my own lights. To take something into account in one's deliberations about what to do is to be motivated by it. Therefore, if one possesses moral understanding then either one is motivated by one's judgements about moral reasons, or one is practically irrational. Therefore, moral internalism is true. The motivation is internal to the reason judgement because judging that the consideration is a reason is enough to get a rational person to take it into account in their deliberations. No external motivational factor is required.

As discussed in chapter two, according to the internalist account under consideration, overall moral judgements are not themselves motivating. When one concludes that overall morally one ought to  $\Phi$ , one does not thereby judge that one has an *extra* reason to  $\Phi$ . The conclusion that one morally ought to  $\Phi$  is not something that provides a new input for moral deliberation; it is a summary verdict on all the morally relevant considerations that have been taken into account in deliberation. However, judging that something is morally right is not necessarily the termination point of deliberation. Moral considerations are not the only considerations that one might take into account. Of course, when one concludes that something is what one has most reason to do, the work of deliberation is done; one only has to actually go on and do it. Overall judgements about reasons can therefore be considered intrinsically motivating, since rational agents are motivated to act in accordance with them. If moral reasons were overriding then overall moral judgements would be like judgements of what one has most reason to do in this respect (at least, they would in rational agents with moral understanding). However, given that we are not thinking of moral reasons as overriding we should not think of overall moral judgements like this. There is no clear sense, therefore, in which we can say that overall moral judgements are intrinsically motivating. The most we can say is that they are indicative of the recognition of moral reasons, and such recognition is intrinsically motivating for rational agents who possess moral understanding. That the recognition of moral considerations is *intrinsically* motivating in this way is sufficient to make it appropriate that the resultant position be deemed a form of moral internalism. Moreover, this position affirms the practicality of morality, hence deflecting the most forceful objection to moral cognitivism.

In chapter two (§ 2.3) we noted the mistaken internalist assumption, attributed by Foot to Kant, that actions must either be motivated by respect for the moral law or by an ulterior motive. Having vindicated moral internalism, we can see a way of interpreting Kant that does not involve attributing to him this oversight. Kant's point is that merely acting in accordance with what morality requires does not make one morally praiseworthy since action from selfish, prudential or even evil motivations might sometimes coincide with what morality requires. We can maintain this insight by saying that an action is morally praiseworthy only if the agent gives appropriate weight to moral considerations. Hence, we do not need to maintain that the moral law needs to explicitly motivate actions in order for them to be praiseworthy. Most probably the moral law and the categorical imperative will rarely, if ever, actually feature in the motivations of morally praiseworthy actions. Their normal role is in reflection about what may be taken to be a reason for action.

This is consistent with Kant's account. Kant thinks of actions as being performed on maxims. Maxims are intentions that relate actions and reasons. What is important for Kant is that actions are performed on maxims that pass the categorical imperative test, maxims that embody good reasons for action. However, we do not need to think of actions as being performed on a single maxim. Rather, deliberation typically involves taking many considerations into account. Anything that one might take to be a reason in favour of performing an action must be, when considered in its relation to the action, capable of passing the categorical imperative test. If it fails this test then it cannot be considered to favour action. It might be thought that since the categorical imperative tests whether a given consideration is, by itself, a sufficient reason for action, it does not test for contributory reasons. Contributory reasons by their nature only *contribute* to overall verdicts. However, nothing can be considered a reason for action unless were it to be the only reason in play (together with an enablers that it requires) it would be sufficient to justify that action, making it overall the act that we should perform. Any consideration that lacks this feature is not a reason for action (although it may be able to enable or enhance other reasons). Therefore, the categorical imperative test yields contributory moral reasons. Since there are moral reasons, moral internalism is vindicated.

One final point about moral reasons requires clarification. I claimed (§ 1.3) that moral considerations are not overriding, and that weak moral obligations might be overruled by weightier non-moral reasons. It might seem that on my account any

consideration, C, that could potentially override an overall moral verdict must itself be considered a moral reason since, in order for it to be permissible to act on C, action on C must pass the categorical imperative test. It might be thought that if C passes the categorical imperative test then it thereby becomes a moral consideration. However, I don't see why we should accept this. Even though the categorical imperative generates specifically moral obligations when maxims fail the test, considerations that pass the test might be either moral or non-moral. Moreover, although there is an overriding reason not to act on a maxim that cannot be willed as a universal law, an overriding reason not to act on a particular maxim is not itself an overriding reason to perform (or not perform) any particular action. It is an overriding reason not to perform a particular action on the basis of a particular consideration. It is therefore not directly a reason for action at all and *a fortiori* not an overriding reason for action. Morality does not, then, provide us with overriding reasons for action. It provides important, but contributory, reasons.

## 8.10 Conclusions

In this chapter I have argued that the shared and universal nature of reasons guarantees that there are moral reasons, and that moral normativity is generated, as Kant claimed, through the rational authority of both the categorical imperative and the duty to always treat rational agents as ends in themselves. I have explained how this relates to the practical reason approach set out in chapter one and argued that the existence of moral reasons vindicates moral internalism. In the final chapter I defend my account against objections and explain how it relates to the possibility of moral knowledge and the problem of moral disagreement.

## Chapter 9. Objections and implications

### 9.0 Introduction

I have argued that there are moral reasons and that therefore we should accept moral internalism. It was claimed in chapter one that, on the practical reason approach, showing that there are moral reasons would also vindicate moral cognitivism, thus demonstrating that moral knowledge is possible. If this is right then the theoretical and practical aspects of morality will be reconciled and moral cognitivism will be immune to the most prominent non-cognitivist attack, the non-cognitivist claim that cognitivism cannot make sense of the practicality of morality. In this chapter I defend my position against a number of potential objections. I then show how my arguments support moral cognitivism. I conclude by returning to the issue of moral disagreement to show how my account meets the other main challenge to cognitivism.

### 9.1 Williams' objections

In *Ethics and the Limits of Philosophy* Williams criticises the Kantian attempt to account for morality “in terms of laws laid down by practical reason for itself” (1985 p. 55). Williams reconstructs the Kantian argument in two different ways and points out what he takes to be fatal flaws with each. He also brings two criticisms against the resultant Kantian position. I start with the objections to the Kantian argument and then consider the objections to the Kantian position.

Williams' first reconstruction of the Kantian argument begins from the desire for freedom. Williams supposes that the Kantian argument depends upon the claim that our desire for freedom entails that we must claim a *right* to our basic freedom. Since we claim this right just on the basis of being rational agents, we must admit that other rational agents also have this right. Williams admits that the last step in this argument is sound:

It rests on the weakest and least contestable version of a ‘principle of universalizability’ which is brought into play simply by *because* or *in virtue of*. If a particular consideration is really enough to establish a conclusion in my case, then it is enough to establish it in anyone's case. (1985 p. 60)

Williams agrees that as rational agents we want freedom, but does not think this commits us to thinking that our freedom is good; he claims that neither wanting something nor deciding deliberately to pursue it commits us to thinking of that thing as good. Williams says, “[t]he idea of something’s being good imports an idea... of a perspective in which it can be acknowledged by more than one agent as good” (1985 p. 58). He thinks that even for someone to think that achieving something would be good *for him* “implies a perspective that goes somewhere beyond the agent’s immediate wants, to his longer-term interests or well-being” (p. 59). Moreover, Williams claims that it is possible for someone to prescribe that others respect their freedom, and to think this reasonable, whilst not prescribing of themselves that they respect the freedom of others. Williams admits that if I must prescribe a rule then I must prescribe that I not interfere with the freedom of others. However, Williams asks, “why must I prescribe any rule?” (1985 p. 62) Why, that is, must I think of myself as a legislator and citizen of a kingdom of ends?

We can respond to Williams by pointing out that I must see others as free and autonomous persons capable of acting for reasons, and I must will that others not treat me as a mere means. My expectation that others not treat me as a mere means is based simply on the fact that I am someone, an autonomous agent capable of reflection and deliberation, capable of thinking in terms of reasons. Since I must act for reasons, and reasons are shared, in acting I necessarily become a legislator<sup>1</sup> and citizen in a kingdom of ends, endorsing the consideration on which I act as genuinely favouring my action. Williams objects:

We are concerned with what any given person, however powerful or effective he may be, should reasonably do as a rational agent, and this is not the same thing as what he would reasonably do if he were a rational agent *and no more*. Indeed, that equation is unintelligible, since there is no way of being a rational agent and no more. (1985 p. 63)

This misunderstands what is at issue. Even if there are many ways of being a rational agent, and no way of being a rational agent and no more, all the ways of being a rational agent might have something in common that guarantees that some ways of acting are

---

<sup>1</sup> I do not become a legislator in the sense that I can determine the content of the law without constraint. The content of the law is supplied by practical reason so in possessing practical reason I provide the content of the law. Therefore, I legislate but without being able to control the content of the law at will. I am also a legislator in the sense that in recognising and endorsing the law (as I must, since I possess practical reason,) I ratify it.

not permissible for us. The Kantian argues that all ways of being a rational agent do have something in common – the necessity of acting on reasons. This takes us to the second way of trying to make the Kantian strategy work that Williams discusses. Williams writes:

If [a rational agent] acts *on* reasons, then he must not only be an agent but reflect on himself as an agent, and this involves his seeing himself as one agent among others. So he stands back from his own desires and interests, and sees them from a standpoint that is not that *of* his desires and interests. Nor is it the standpoint of anyone else's desires and interests. That is the standpoint of impartiality. So it is appropriate for the rational agent, with his aspiration to be genuinely free and rational, to see himself as making rules that will harmonize the interests of all rational agents. (1985 pp. 65-6)

Williams thinks that this adequately characterises the theoretical standpoint in which we decide what to believe, but not the practical standpoint in which we decide what to do. This is because in asking what I should believe I ask what *anyone* should believe given the evidence at hand. Williams claims that the practical standpoint, unlike the theoretical standpoint, is essentially first-personal, since “[t]he *I* that stands back in rational reflection from my desires is still the *I* that has those desires” (1985 p.69). According to Williams, theoretical reflection is impartial because we are concerned to discover the *truth*. Practical reflection does not have this goal, so if it is impartial it must be for another reason. It would only be impartial, Williams claims, if there is a requirement for us to be committed to equality and harmonising the interests and outcomes of deliberation of all rational agents. However, Williams claims, the assumption that there is such a requirement has received no support from the Kantian argument.

The force of Williams' objection depends upon exaggerating the difference between the theoretical and the practical. Seeking the truth does indeed seem to involve impartiality because there is a single world that we are trying to get things right about. However, theoretical reflection is also first-personal in the way that Williams attributes to practical reflection. In reflecting on what I am to believe I do not discover a God's eye perspective; I must begin from where I am – my own perceptions and beliefs. The *I* that stands back in rational reflection from my beliefs and perceptions is still the *I* that

has those beliefs and perceptions<sup>2</sup>. Moreover, Williams' contrast between the theoretical and the practical seems already to presuppose that in practical reflection, unlike theoretical reflection, we are not concerned with trying to get things right. This presupposition begs the question against the Kantian who maintains that there *are* distinctively practical ways of getting things right, and these are embodied in the categorical imperative and the moral law.

As well as objecting to the Kantian argument, Williams objects to the resultant Kantian position, complaining that, “[i]f morality is possible at all, does it leave anyone in particular for me to be?” (1985 pp. 69-70) However, on the Kantian account we can only develop integrity, an identity and a personality, if we are committed to taking some considerations as reasons – committed to not acting whimsically. Morality arises from taking practical reason seriously, and it is only this that enables us to become someone in particular. Even when morality demands sacrifices of us, its demands are the demands of particular personal relationships, of treating others as persons. The requirements of morality are not the requirements of an impersonal system but the requirements that arise from our obligations to others. Since it is only through our relationships to others that we can become someone in particular, that we can become anyone at all, morality supports, rather than undermines, our integrity and identity.

Williams' other objection is that Kantian accounts, and all systems of morality more generally, fall into error by trying “to make everything into obligations” (1985 p. 180). Williams writes:

Suppose you are under an everyday obligation—to visit a friend... because you have promised to. You are then presented with a unique opportunity, at a conflicting time and place, to further significantly some important cause... You may reasonably conclude that you should take the opportunity to further the cause. But obligations have a moral stringency, which means that breaking them attracts blame. The only thing that can be counted on to cancel this, within the economy of morality, is that the rival action should represent another and more stringent obligation. Morality encourages the idea, *only an obligation can beat an obligation* (*ibid.*).

This is puzzling, Williams thinks, because intuitively the pursuit of an important cause can overrule an obligation even when there is no obligation for us to pursue the cause in question. I have already partially addressed this worry by giving an account of moral

---

<sup>2</sup> We should not overlook the fact that in both practical and theoretical reflection I am unlikely to start simply from my own individual perspective; I am likely to listen to and reflect on what others might have



reasons as contributory (§ 8.9). Specifically, as far as the categorical imperative test is concerned, I think we can answer Williams' objection by thinking of factors such as important causes as entering into the background conditions against which maxims are to be tested. Thus, the fact that I can further some important cause can allow my maxim to pass the test when it would otherwise have failed. I can be released from an *overall* moral obligation to keep a promise, even though I still have a contributory moral reason to keep my promise, by the fact that the maxim of promise-breaking-when-an-important-cause-can-be-significantly-furthered can pass the categorical imperative test. When a maxim passes the test no obligation is generated, so in a sense important causes *can* 'defeat' obligations. Strictly speaking they do not defeat obligations however: they mean that an obligation is never generated in the first place.

## 9.2 MacIntyre's criticism of Kant

MacIntyre claims that the test of whether maxims can be consistently willed does not succeed in distinguishing between morally required and forbidden maxims. He raises familiar counter-examples of two kinds – maxims that pass the test that are trivial and non-moral, and those that pass that, according to commonsense morality, are immoral. MacIntyre ridicules Kant's claim that "any man who wills the maxim 'To kill myself when the prospects of pain outweigh those of happiness' is inconsistent because such willing 'contradicts' an impulse to life implanted in us all" (MacIntyre 1987 p. 45), writing, "[t]his is as if someone were to assert that any man who wills the maxim 'Always to keep my hair short' is inconsistent because such willing 'contradicts' an impulse to the growth of hair implanted in all of us" (*ibid.*).

Admittedly Kant's argument (1785 § 422) that suicide is impermissible is at best unclear. However, MacIntyre's criticism indicates that he misunderstands Kant's views: 'always to keep my hair short' is not a maxim at all, since it mentions<sup>3</sup> no purpose or reason for which the action is performed. If killing myself to avoid pain is impermissible it is because of the relation between the action and its reason or goal, not because of the nature of the action alone. Plausibly, killing myself to avoid pain is

---

to say.

<sup>3</sup> It might be thought that maxims do not need to *mention* purposes, even though they *presuppose* them. However, for the Kantian, the permissibility of action on a particular maxim is dependent upon the relation between the action and the reason or purpose. Kantians do not claim that particular act types (e.g. telling of untruths) are intrinsically impermissible, but that acts of particular types can only be justified by

impermissible because in performing the action one aims to destroy the person for whose sake the action is performed. To treat someone as an end is to treat them as a reflective being who has reasons for action that reflect their reasoned commitment to certain ends, projects and purposes. Perhaps I cannot treat myself as an end if the purpose of my action is to ensure that I no longer exist as a person. However, it is not necessary to decide this question here. The important point is that despite the weakness of Kant's own discussion of suicide, MacIntyre's objection does not highlight any problem with Kant's central claims.

Let us move onto the objection that trivial and non-moral maxims can pass the categorical imperative test. MacIntyre gives the following examples:

‘Keep all your promises throughout your entire life except one’, ‘Persecute all those who hold false religious beliefs’ and ‘Always eat mussels on Mondays in March’ will all pass Kant's test, for all can be consistently universalized. (1987 pp. 45-6)

Again, the imperatives that MacIntyre lists are not maxims since they do not relate an action to a reason or purpose. Just proposing a policy of acting in a particular way is not a rational way of determining one's conduct. One must do so *because* one sees that there is something to be said for it – some reason for it. Secondly, it is not problematic for Kantians if trivial non-moral maxims can pass the categorical imperative test. When a maxim passes the test it simply means that it is permissible to act upon it. This does not entail that it is a maxim that we are *morally required* to act upon. Thirdly, for Kantians morality is to be understood not just in terms of the categorical imperative test but also in terms of the requirement never to treat another as a mere means. Regardless of whether these two aspects of morality are equivalent or complementary, it is unfair to criticise the Kantian position without taking both into account.

Fourthly, of the three principles that MacIntyre lists, the first two (the supposedly immoral ones) are unlikely to pass the categorical imperative test when they are formulated into maxims. It seems there can be no legitimate reason for wanting to keep all of one's promise except one, and acting on this maxim seems to make promise keeping or breaking a mere means to completing some idiosyncratic project. Since by extension this involves treating others as a mere means this is morally unacceptable on Kant's view. The case of religious persecution is more complicated, since the example

---

particular reasons or purposes. To assess the permissibility of a maxim it is therefore *essential* that the

lacks the absurdity of the promise-keeping example. There are a variety of factors that can and do motivate religious persecution. Can any of these feature in a maxim that can pass the categorical imperative test? Perhaps the most likely candidates are cases where someone persecutes those of other religions in order to eliminate and prevent the spread of perceived falsehoods or in order to eliminate ways of life that they consider immoral. Hence we can derive the following maxims:

- Maxim 1: I will persecute anyone who holds false religious beliefs in order to prevent the spread of those false religious beliefs.
- Maxim 2: I will persecute anyone who holds false religious beliefs in order to eliminate ways of life that are immoral.

It might be thought that these maxims are permissible on the Kantian scheme because:

- (a) Plausibly everyone can agree to the goals of eliminating immorality and preventing the spread of falsehoods.
- (b) The acts of persecution do not treat their victims as a mere means because they aim to get their victims to abandon false beliefs and immoral practices.

There are two problems with this. Firstly, it seems highly unlikely that there is any basis on which someone can consent to acts of persecution. Persecution by its nature involves hostility towards the values, practices and identity of its victims<sup>4</sup>. It involves attempting to destroy those values and practices, to destroy that identity. No one can consent to action that aims to destroy their own values, practices and identity. Secondly, consider the case of someone who, believing Islam to be a false religion, is considering persecuting Muslims to try to prevent the spread of Islam. They act on maxim 1. Even if a Muslim can consent to action on maxim 1, they cannot consent to persecution of Muslims that is based on maxim 1 because necessarily any Muslim, believing Islam to

---

reason or purpose for the action be explicitly mentioned.

<sup>4</sup> If we replace the word 'persecute' with 'attempt to suppress' then it may not follow that the actions in question are necessarily impermissible in the way described. However, this all depends upon the means of suppression chosen and whether they involve respect for persons rather than using others as a mere means in a drive to stamp out false beliefs. It is therefore unlikely that the permissibility of such suppression will constitute a counterexample to the Kantian picture.

be true as they do, believes that maxim 1 should not lead anyone to persecute Muslims<sup>5</sup>. Since no one believes their own religious beliefs to be false no one can consent to being persecuted on the basis of maxim 1. Acting on maxim 1 is therefore impermissible. It involves treating others as a mere means. Analogous considerations apply to action on maxim 2. Since no one believes their own religious practices to be immoral they cannot consent to action that aims to eliminate the immoral practices of their religion.

MacIntyre's objections to the categorical imperative are inconclusive at best. However, he also objects to the principle that we ought never to treat others as a mere means, claiming that Kant has given us no reason to adopt this principle. MacIntyre writes:

‘Let everyone except me be treated as a means’ may be immoral, but it is not inconsistent and there is not even any inconsistency in willing a universe of egoists all of whom live by this maxim (1987 p. 46)

If MacIntyre means to point out that there is nothing incoherent about a world populated exclusively by egoists then he is surely right. However, *that the world should be populated exclusively by egoists* is not something that *I*, as one of the agents in the world, can will. I cannot will that others treat me as a mere means. Being treated as a mere means involves having one's will disregarded. I cannot coherently will that my wish to have my will disregarded itself be disregarded, since I then will that others treat me as an end whilst simultaneously treating me as a mere means. Since to adopt an end is to will it, I necessarily will the achievement of my own ends. I therefore cannot will that my own ends be frustrated since willing that my own ends be frustrated is to will that what I will does not come about. In this case, I no longer will it as an end. If I do not will that I am treated as an end then I do not will at all. Therefore, I cannot will a universe of egoists who all will that everyone except themselves be treated as a mere means.

---

<sup>5</sup> A revised maxim might be proposed to avoid this: ‘I will persecute anyone who holds religious beliefs that I believe to be false, in order to prevent the spread of false religious beliefs’. Now there is another reason why the maxim cannot be universalised. In order to will the maxim as a universal law I need to will that even people with true religious beliefs are persecuted: by those who believe their (in fact true) religious beliefs to be false. Moreover, I must will that I am persecuted by anyone who disagrees with my religious beliefs, and that by doing so they will prevent the spread of false religious beliefs. Over and above the difficulty of willing my own persecution, I must also will that my false beliefs be reformed. Since I necessarily do not believe my religious beliefs to be false, I cannot will this.

### 9.3 Cohen's objection to Korsgaard

Gerald Cohen notes two simple arguments that appear in Hobbes' *Leviathan*. Firstly:

[I]f you make a law, then that law binds you, *because* you made it. For if you will the law, then how can you deny that it binds you, without contradicting your own will? (Cohen 1996 p. 167)

Secondly:

[I]f you are the author of the law, then it *cannot* bind you. For how can it have authority over you when you have authority over it? How can it *bind* you when you, the *lawmaker*, can change it, at will, whenever you like? (Cohen 1996 p. 167)

Cohen points out that this claim is not quite right, since it is based upon the false assumption that "if I *can* repeal the law, then it fails to bind me even when I have not *yet* repealed it" (p. 170). Nonetheless, a law that I can repeal in one sense does lack categorical authority over me, since whenever this law commands something of me, I can view the imperative as having a clause: 'you must either  $\Phi$ , because the law requires it, or you must repeal the law'.

Cohen says that on Korsgaard's view "the subject of the law is also its author: and that is the ground of the subject's obligation – that *it* is the author of the law that obliges it" (Cohen 1996 p. 170). In contrast, "Kant thought that if the moral law came *just* from my own will, then it would have no claim on me, rather as the law of the sovereign has none over the sovereign in Hobbes" (1996 p. 171). Cohen says that Kant thought that a person "indeed makes the moral law, but that he cannot unmake it, for he is designed by nature to make it as he does, and what he is designed to make has the inherent authority of reason as such" (*ibid.*). Hence, Kant thought that although we are legislators of the kingdom of ends as well as citizens of it, we are not sovereigns. The consequence of this difference between Korsgaard's position and Kant's is that, according to Cohen, Korsgaard may not be able to answer the question of why, if I am the lawmaker, I cannot repeal the law that I make whenever I please. Cohen writes:

Suppose I ask: *why* should I obey myself? Whom am *I*, anyway to issue a command to *me*? Kant can answer that question. He can say that, although you *legislate* the law, the content of the law comes from reason, not from anything special about you, or *your* reason, or even *human* reason, but from reason as

such. And, when that is so, then, perhaps, reflective endorsement of the law is inescapable. But if the content of the law reflects *my* nature, my engagements and commitments, then could I not *change* its content? (1996 pp. 173-4)

Korsgaard emphasises the role of practical identity, and sees individuals as possessing a number of practical identities that ground reasons and obligations for them. The most fundamental of these is their nature as a reflective being, but their other identities, such as their identification with a particular ethnicity, nation, religion, football team or political party can be changed or at least stripped of their normative significance. However, even if Korsgaard is right that these other contingent practical identities are normative, and that they make laws for us that we can repeal by rejecting those practical identities, it does not follow that the laws that practical reason lays down for itself are ones that we can repeal. If practical reason lays down laws for itself, they are inescapable laws for any being capable of deliberative and reflective action. With respect to the moral law, we are citizens and legislators, but not sovereigns. Reasons are shared and action on a reason cannot be whimsical. As an individual, I do not have the authority to decide that some consideration is or is not a reason. That authority comes from reason, and is shared amongst all the members of the kingdom of ends – among all creatures that possess reason.

#### 9.4 Wallace's objection

Wallace formulates a dilemma for Korsgaard's position. Wallace says that on Korsgaard's account, "[t]he unitary source of normativity... is the commitment to comply with such normative principles as the instrumental principle and the categorical imperative, a commitment that is in turn built into all concrete determinations of the will" (2006a p. 78). Wallace suggests that there are two ways of interpreting this claim, and neither is plausible. Firstly, we might think of normativity as having its source in "brute motivational tendencies... to comply with the principles whose normativity is in question" (p. 79). However, normativity can only be generated from an inclination or motivational tendency if we already have reason to endorse or follow our inclinations or tendencies. Since a prior normative reason is required, motivational tendencies cannot be the source of normativity. Secondly, "we might try to interpret such commitments not as brute motivational tendencies, but rather in essentially normative terms" (*ibid.*). However, to avoid suggesting that moral considerations are normative independently of,

and prior to, our volitional relation to them, we must locate the normativity *in* our commitment to act on those principles. Instead of seeing agents as having a brute motivational tendency to act in accordance with (e.g.) the categorical imperative we will think of them as having such a tendency because they are committed to seeing the categorical imperative as normative. However, Wallace objects, this mislocates normativity; it does not allow us to conclude that the categorical imperative is actually normative. Wallace writes:

From the fact that I am, say, committed to the normative principle that it would be good to thwart my enemy, it cannot be inferred that it really would be good for me to thwart my enemy (*ibid.*).

Neither a brute commitment to a principle nor a commitment to the normativity of a principle can make that principle normative. Wallace notes that Korsgaard's argument might instead be given a transcendental interpretation. If *we* are committed to the normativity of (e.g.) the categorical imperative simply by being deliberating agents then there is no position from which we can intelligibly challenge the normativity of the categorical imperative. Since we cannot cease to be deliberating agents, we cannot cease to see the categorical imperative as normative and we cannot escape our commitment to the categorical imperative. Wallace suggests that this move only appears effective because terms such as 'commitment' conceal an equivocation. It is one thing to believe that it would be good to act in accordance with a principle and another to be motivated to act in accordance with it.

The practical reason approach can help us to overcome this problem. Whilst believing that one has a reason,  $r$ , to  $\Phi$  is not identical to being motivated to  $\Phi$ , to being committed to action on the basis of  $\Phi$  or even to taking  $r$  into account in deliberations on occasions where one might  $\Phi$ , there remains a close connection between cognition and action here. Specifically, if one believes that  $r$  is a reason to  $\Phi$  and yet takes no account of this in deliberating about whether to  $\Phi$  then one is irrational, even by one's own lights. Thinking that a particular consideration speaks in favour of a particular action is rationally incompatible with not taking it into account, since considerations that favour are considerations to be taken into account. The practical reason approach allows us to overcome this problem since it is in practical reasons that the practicality and objectivity of morality come together. This means that if I see myself as having a reason to conform to the categorical imperative then I am thereby motivated to conform

to it (unless I am irrational or think that I have better reasons not to conform to it). We can understand the Kantian argument as showing that we have a reason to conform to the categorical imperative: only by so doing can we act for good reasons.

Despite the fact that this response is available to us, I think that Wallace's objection gets something right. I do not think that we should see Korsgaard as showing that all normativity is generated by volitional commitment and that there would be no normativity without such commitment. I think that we should not attempt to say anything about whether or not normativity would exist in a world that lacked agents. We should not claim that reasons are not real or that they are generated only by our reflection, deliberation and volition. We do not need to deny that there are normative truths. We do not need to claim either that there is a single source for normativity, or to deny that normative considerations may be diverse. We can see Korsgaard's arguments as answering our doubts about whether morality is normative, rather than seeing them as locating the universal source of normativity. Moreover, even though Korsgaard is supposedly a constructivist, and Wallace a normative moral realist, their positions might even be compatible.

Korsgaard (1996b p. 108) suggests that moral principles have normativity built into them. Good maxims are normative in virtue of the relation of their parts. Similarly, in explaining normative moral realism, Wallace writes:

The normativity or bindingness is built into the content of the principle, its standing as a principle that specifies what we have reason to do. The question whether we are obligated to comply with a given principle is already settled by whatever considerations establish it, in the first place, as an expression of the independent truth of the matter about what there is reason to do. (2006a pp. 75-6)

Wallace takes himself to be contrasting his position with Korsgaard's position. However, if we move away from thinking of substantive and procedural realism as incompatible alternatives we can think of normativity as independent of the will whilst maintaining that our doubts about the normativity of morality cannot be answered by merely insisting that there are independent normative moral facts. To answer such doubts we need to begin from something that retains a practical grip on us even when we fall into doubt. In showing that morality is practically necessary for us just because we are reflective agents we overcome our doubts but we do not thereby show that there



are not *really* any normative moral principles or that moral principles are normative only because of their relation to the will.

Thinking in this way cuts across a common way of distinguishing between Kantian constructivist positions and realist positions. Shafer-Landau, for example, writes:

For the Kantian, there is no moral reality... if there is no such thing as pure practical reason... For constructivists, what reality there is depends crucially on the existence of the preferred standpoint... Realists believe that there are moral truths that obtain independently of any preferred perspective, in the sense that the moral standards that fix the moral facts are not made true by virtue of their ratification from within any given actual or hypothetical perspective. That a person takes a particular attitude toward a putative moral standard is not what makes that standard correct. (2003 pp. 14-5)

However, on the approach that I have been advocating we cannot really distinguish between these two positions. Upon reflecting on our own nature as reflective agents we see that moral principles are necessarily intrinsically normative. We cannot step outside of this perspective to assert or deny that moral normativity really would – or would not – exist independently of our reflective nature. Even though when we fall into doubt it is our reflections on our rational nature and agency that allow us to reaffirm moral normativity, we cannot think of moral normativity as being constructed from these reflections or their contents. Kantians should therefore affirm the reality of the normative.

## 9.5 Enoch's objection

Enoch asks us to imagine that we manage to convince a sceptic who is indifferent to morality that “you cannot even count as an agent and your bodily movements cannot even count as actions unless you aim at self-constitution of the kind Korsgaard has in mind” (2006 p. 179). Enoch suggests that the sceptic might then ask why it should matter whether he counts as an agent, why he should care about whether he performs actions. By analogy, if we tell a shoddy builder that he is not living up to standards that are constitutive of building a house, he might ask why it matters that he builds a house. Why should he not adopt his own, related project of building a ‘schmouse’? Likewise, the sceptic who fails to live up to standards constitutive of agency might nonetheless

live up to standards of ‘schmagency’. Why, the sceptic might ask, should it matter whether we classify him as an agent or a schmagent?

It seems unclear, therefore, whether the fact that acting in accordance with a standard is constitutive of agency is relevant to the normative credentials of that standard. Enoch considers three potential replies to this worry drawn from Rosati (2003 p. 522). The first is that “being constitutive of agency renders the relevant motives and capacities nonarbitrary, thus normatively vindicating them” (Enoch 2006 p. 181). It is true that being constitutive of agency seems to make a standard non-arbitrary, since it picks it out as special and interesting in some respect. However, it is unclear whether being non-arbitrary in this sense is sufficient for normative vindication. The standards that are constitutive of chess playing are also non-arbitrary (they have been designed and adjusted in order to make chess the kind of game that it is). This does not show us that they are normatively binding on all rational agents. There may be good reasons for playing chess, but nonetheless chess playing is a rationally optional activity. There is no reason why we shouldn’t invent and play a related game, schmess, with a different set of rules. If being constitutive of a practice is sufficient for being non-arbitrary then the rules of schmess and schmagency are non-arbitrary. If, additionally, being non-arbitrary is sufficient for normativity then the rules of schmess and schmagency are normative. It would follow that we all have reason to follow the rules of agency and schmagency and to be both agents and schmagents. The problem, then, with this response to the challenge is that it merely asserts that being constitutive of agency is sufficient for normativity. What is required is an argument that shows *why* this is so. We need to know what is special about agency that allows it to ground normative requirements in ways that other practices or capacities do not.

A second reply to the challenge is to claim that our conformance to standards constitutive of agency is normatively vindicated by the fact that we care about conforming to these standards and this caring “involves no identifiable mistake” (Rosati 2003 p. 522). Enoch (2006 p. 182) highlights two problems with this response. Firstly, it is unclear that we do actually care about living up to the standards that are constitutive of agency. Secondly, the criterion of not involving an identifiable mistake fails to distinguish between caring about the standards that are constitutive of agency and caring about other things such as pleasures and pains. This criterion therefore cannot support the Kantian claim that the former, but not the latter, are directly normative.

The third possible response to the worry is that “there is something self-defeating about challenges to the capacities and motives constitutive of agency, and so they have a self-vindicating status” (Enoch 2006 p. 181). Enoch invites us to consider sceptical challenges in other contexts, such as the sceptical attack on logic. Nagel (1997), for example, argues that since sceptical attacks on logic must themselves use logic, they are self-defeating or, at least, unstable. Enoch rejects the claim that our use of logic is therefore self-vindicating. He claims that sceptical challenges often involve “highlighting tensions within our own commitments” (2006 p. 183), adopting a structure similar to a *reductio ad absurdum* that entitles them to begin by presupposing the commitments of the position they aim to undermine. Even if the moral sceptic relies on the capacities that constitute agency in mounting their sceptical challenge this in itself does not show that these capacities are normative<sup>6</sup>. The claim that the sceptical challenge is self-defeating is therefore unconvincing.

Enoch considers whether the fact that we cannot opt out of being agents is able to help us in grounding the normative status of moral requirements in agency. However, the sceptical challenge re-emerges in this context unless we are able to show that we have reason not to be half-hearted in pursuing the aims constitutive of agency. Enoch writes:

Think again about finding yourself playing a game of chess, and assume ... that you cannot quit. And assume that sacrificing a pawn is the thing you have most chess-related reason to do (it best promotes your chances of checkmating your opponent or some such). Well, do you have a reason to sacrifice a pawn? Not, it seems to me, if you don't have a normative reason to play or win the game, and this even if you can't quit. (2006 p. 189)

Enoch's suggestion is that unless I have a reason to be an agent then I have no reason to do anything that is constitutive of being an agent<sup>7</sup>. Let us consider this worry in the light of normative nihilism. Normative nihilism is the claim that there are no reasons for action, no reasons at all. Suppose that normative nihilism would be true unless there are reasons for action that spring from the standards that are constitutive of agency. On this

---

<sup>6</sup> Korsgaard (1996b p. 15) herself recognises this point in acknowledging that even if we were to discover that moral requirements are not normative and we gave up our endorsement of them we might nonetheless find ourselves unable to help acting morally.

<sup>7</sup> The idea of needing to have a reason to be an agent sounds strange, but I overlook this strangeness in order to consider the full force of Enoch's argument. Just as, if forced to play a game of chess, one can play it half-heartedly or wholeheartedly, it might be possible for us, even though agency is inescapable for us, to play the game of agency half-heartedly or wholeheartedly. If so, we may need a reason to endorse our status as agents and to excel at agency.

assumption there cannot be any reason to be an agent. Since all reasons are supposed to spring from adopting the standards that are constitutive of agency, there can be no prior reason to adopt those standards. Then, if Enoch is right, there can be no reasons at all. Suppose instead that we reject normative nihilism but we are unsure whether there are any *moral* reasons. Now we think that we have reasons to do lots of things. Necessarily one of the things we have reason to do is to choose on the basis of the normative reasons that we have – for to think that something is a reason is to think that it favours action. Since to be an agent is to exercise the capacities for reflective self-determination, deliberation and choice on the basis of reasons, we have reasons to be an agent. Hence, if there are any reasons for action there are reasons to be an agent. Since we are reflective beings who must make practical choices to act on the basis of considerations that we take to favour action, normative nihilism is not an option for us from a first-personal, deliberative, perspective. Since we are committed to acting on the basis of reasons, we are committed to agency and to there being reasons for being an agent.

It seems that Enoch is right to claim that “normativity cannot be grounded in what is constitutive of agency” (2006 p. 192). Nonetheless, it might be possible to ground *moral* normativity in agency together with the assumption that there are at least some other considerations that are normative for us. Since a sceptic who claims that there are no reasons for action at all must be being disingenuous (they have to act, after all), this tactic, if successful, will convince the sceptic. Enoch writes:

I think Korsgaard is right in emphasizing the unavoidability (for us) of deliberation or reflection, and that she is also right in taking reasons or other normative facts to be needed for deliberation. But I think she is mistaken in thinking that any of this counts in favor of grounding normativity in what is constitutive of agency or of action. (2006 p. 194)

However, once the sceptic has accepted that deliberation is unavoidable for us as rational agents and that reasons are shared and universal, they cannot therefore avoid the conclusion that *either* it is permissible for them to be treated as a mere means, or they must treat others as ends in themselves. Since deciding to act for a reason on the basis of reflection and deliberation *is* to constitute oneself as an end, no one can view themselves as a mere means. Even if I were, for example, to devote myself exclusively to making others happy, I would not thereby be able to see myself as a mere means to their happiness. In pursuing the happiness of others I constitute myself as an end by choosing on the basis of reasons, acting autonomously and pursuing an end that I have

reflectively decided to be worthwhile. This is not to say that Enoch is wrong. Enoch's point is that if we assume that there are no irreducibly normative truths then we cannot generate normativity from that which is constitutive of agency. As Enoch says, "our best attempts at deciding what to do do require irreducibly normative facts" (2006 p. 195). This gives us a good reason to believe in normative facts, just as the fact that "our best attempts at making sense of the world ... require the existence of electrons is ... [a] reason to believe in the existence of electrons" (*ibid.*). If we commit to there being reasons for action (as, indeed, we must), then our reflections lead us to see that moral reasons must be among their number.

## 9.6 The objection from particularism

There is an apparent objection to the position I am advocating that comes from within the practical reason approach. It begins from reflections about the nature of reasons and concludes that moral principles should play no important role in moral philosophy and are unnecessary for, or even a hindrance to, living a moral life. The resultant position is moral particularism, which is the following claim:

Moral thought, moral judgement, and the possibility of moral distinctions – none of these depends in any way on the provision of a suitable supply of moral principles (Dancy 2004 p. 5; cf. McKeever and Ridge 2006 p. 19).

Since Kantian positions are typically taken to be paradigmatically principled systems of morality, it seems that moral particularism, if true, would show that Kantian positions are false. Dancy bases moral particularism on holism about reasons, a thesis that I have been presupposing in my practical reason approach. Therefore, if the argument from holism about reasons to particularism is sound this might undermine my position.

Holism about reasons is the claim that "a feature that is a reason in one case may be no reason at all, or an opposite reason, in another" (Dancy 2004 p. 73). Holism about reasons is to be understood in opposition to atomism about reasons, which is the claim that "a feature that is a reason in one case must remain a reason, and retain the same polarity, in any other" (p.74). It is disputed whether there is a valid argument from holism to particularism (McKeever and Ridge 2006), and even Dancy admits that if there is, it is "at best indirect" (Dancy 2004 p. 82). To avoid these complications, I suppose for the sake of argument that there *is* a valid argument from holism to particularism but that my position is compatible with the truth of particularism. My

position does not depend upon there being any universally valid moral principles, with the exceptions of the categorical imperative and the principle that we ought never to treat others as a mere means. These principles are not preconditions for competence with moral concepts, nor are they necessary to directly guide our actions. Rather, they serve as reflective tools that facilitate the understanding of moral reasons.

If holism entails that moral principles are unnecessary, it does so because it shows that principles such as ‘don’t lie’, ‘keep your promises’, ‘be kind’ and ‘be considerate’ have exceptions that cannot themselves be codified in principles. Sometimes kindness or considerateness might count against an action: mopping the brow of a torturer might be kind or considerate but that is no reason to do it (Dancy 2005 p. 9, cf. Kirchin 2007 p. 19). The moral law and the categorical imperative operate in a different way to this kind of principle, and hence are unaffected by arguments that show that such principles are not a precondition for morality. To threaten the Kantian position, particularists would need to show that treating someone as a mere means, whilst usually a reason against action, can sometimes be no reason at all, or a reason in favour. This is something that they have not done, and arguably cannot (and presumably do not wish to) do. They would also need to show that the fact that a maxim cannot be willed as a universal law does not always show that it should not be acted upon. It is not part of the particularist project to show that this is so.

The final point that needs to be made is that, as I understand them, neither the categorical imperative nor the moral law entails that there are exceptionless moral principles such as ‘don’t lie’ and ‘don’t steal’. The categorical imperative test, for example, shows us that certain considerations fail by themselves to provide any reason for stealing or lying – the prospect of personal gain, for example. We can hold onto this insight whilst admitting that typically situations are complicated, involving a number of considerations along with various enablers and defeaters. There can sometimes be sufficient reasons for lying or stealing and this, I believe, is something that Kantians should accept (even if Kant himself did not accept this). The categorical imperative shows us that some considerations cannot function as reasons for us to perform certain kinds of action, whilst others must. If we accept holism about reasons and the argument from holism to particularism we will add that this does not thereby provide us with universal principles on which to act. Whether particularism is true or not, we can still hold that there are moral reasons, and that the categorical imperative and the moral law show us why we cannot deny the normativity of morality.

## 9.7 Other objections

The account I have been defending might be considered broadly rationalist; it is based upon a practical reason approach that emphasises the importance of reflection, deliberation and deciding on the basis of reasons. There are a number of related potential objections to this approach. It might be thought that it underestimates the role of emotion and empathy in morality. However, although I have said little about empathy and the emotions, I think that these phenomena can play an important role on a practical reason account. Plausibly, being appropriately responsive to many kinds of reason involves the right kind of emotional attunement to the situations that one finds oneself in. Since the emotions of others are often highly morally relevant, providing us with (moral) reasons for action, empathy may often be a precondition for such attunement. There is much to be said about this, and I regret that space does not permit discussing it more fully.

It might also be thought that I have allowed too little room for passions and personal considerations to play a role. I am, it might be claimed, offering a view of human life that involves only impartiality, disinterestedness and the cold machinations of reason. This is untrue. I have not suggested that reasons and values are impersonal, but that they are interpersonal. Moreover, I have not attempted to give a complete account of well being or the good life and I have not suggested that moral considerations are overriding. Even within the moral domain, my position is not necessarily impartialist in the sense of ruling out partiality towards persons and projects with which one has a particular relationship. Furthermore, despite my insistence that our reasons are not provided by desires, I have not claimed that our desires and passions are irrelevant to what we should do. Whilst desires are not automatically reason-giving, our motivated desires give us reasons when they are themselves reasonable; plausibly, desires that we endorse and whose satisfaction is important to us and will bring us happiness provide us with reasons for action.

A further objection comes from the Humean theory of motivation. Even though we have seen decisive reasons to reject the Humean theory, it might remain a threat to Kantians since Humeans might claim that there are phenomena that can be made sense of by the Humean theory but not by the Kantian account. This is Russell's (2004) tactic. He claims that the Humean theory always has a ready explanation of how reasons carry

motivational force. However, on the Kantian account, Russell says, “we are asked to accept that pure practical reasons nevertheless carry motivational power in virtue of their rational ‘authority’” (2004 p. 9) Russell continues:

When the Kantian theorist is challenged to explain how this is possible the relevant reply, Korsgaard argues, is to say simply that rational agents, in so far as they are rational, must be motivated by their reasons. No *source* of motivation has been identified or described except the standing condition of rationality itself (*ibid.*).

If a normative reason motivates someone, perhaps we need to give some account of how it does so. This is, in part, because normative reasons do not always succeed in motivating. Humeans claim that just saying that someone is rational is not such an explanation because it does not identify a specific source for the motivation. Therefore, Humeans claim, a motivational explanation needs to connect to a prior motivation of the agent. Saying that someone is rational identifies no such motivation because being rational does not involve having any particular motivations.

This Humean objection is not valid. From the explanatory point of view there is no need for us to mention a prior motivation of the agent. We saw in chapters four to seven that when an agent acts for a reason, whether it is a good reason or not, this does not entail that the agent had any prior related desire. Moreover, in chapter three we saw that generally explanations of action do not even require mentioning specific psychological states of the agent. Mentioning norms, situational factors, character traits, relationships, habits, customs, personal idiosyncrasies, or ongoing projects are just some of the ways that we can explain actions. Moreover, whether from a third personal explanatory point of view or from a first personal perspective, we can understand actions as having been brought about directly by the agent’s decision on the basis of reasons, by their active will. In this case we do not need to identify a prior desire from which the agent deliberated, from which the decision inherited its motivational ‘power’ and from which “pure practical reasons actually generate ‘motivational force’” (Russell 2004 p. 9). If the agent has an active will, if they have the active power of self-determination, they are not enslaved to their prior motivations. An agent might decide on the basis of evaluative beliefs, on the basis of something cognitive. If we oppose the cognitive to the motivational then we may find this confusing, but we have seen that we have no reason to accept this opposition. Cognition and motivation both involve the



interplay of activity and passivity, spontaneity and receptivity (McDowell 1994 pp. 9-23)

Russell claims that motivation “cannot be created *ex nihilo*; it must draw on some source already present in the agent’s psychological disposition” (2004 p. 9). However, Russell does not explain why this is so. We have no reason to accept this assumption. Russell demands from Kantians an explanation of how pure practical reason can set the human body in motion, of “how it is that pure practical reason acquires *causal traction* in the world” (2004 p. 10). On the Kantian account, we, as agents, cause things to happen, and we do so by regulating our conduct in accordance with reasons. Pure practical reason may be the faculty that allows us to do this, but there is no reason for thinking that no causal explanation will be available of the workings of this faculty. Indeed, there *must* be some explanation of how this occurs; what I have argued in previous chapters is that we have reasons to believe that the Humean theory is not adequate as such an explanation. In chapter eight I affirmed that our faculty of practical reason is the product of induction into the space of reasons, of the development of our second nature, our rational nature. Being *rational animals*, as we are, our *rationality* is part of the *natural* world. When we are doing philosophy, or at any rate, when we are trading in reasons, we do not need to be worried about giving an explanation of *how* our second nature fits into the natural world. If we are committed to the scientific worldview we will be convinced that there is such an explanation; the idea of second nature allows us to see that there is nothing particularly mysterious about the phenomenon of rationality. However, it is unreasonable to demand that we give such an explanation in technical scientific terms, and false to claim that we cannot give such an explanation in ordinary terms: we can, and do, give such explanations in the ways that I have described in chapter three. Humeans may be unhappy that these are not sufficiently psychologistic, but they still owe us some reason for thinking that our ordinary explanations of actions must be in terms of mental states.

There is a final objection that I wish to mention. It might be thought that I have not really defeated moral scepticism, since people are quite free to ignore the demands of the categorical imperative and the moral law. Sceptics who do not care about morality may well continue to be disaffected even after they have read and understood the account that I have offered. This is true, but I do not think that this really constitutes an objection. We cannot force people to care about morality, and perhaps it would be immoral to try. However, we can offer people good reasons for caring about morality

and do what we can to increase moral understanding. Morality is about how we should relate to each other, about having decent, caring and respectful relationships (Korsgaard 1996a pp. 275, 310). Since almost everyone cares significantly about people and relationships, typically people should readily understand, and be motivated by, moral reasons. The account that I have given here should help us to understand why this is so, should strength our moral commitments, and should help us when we fall into doubt. When we encounter those who do not care about other people or relationships, we can try our best to convince them that these things matter. However, if they fail to be convinced by our account or fail to see why it matters this should not make us worry that we have got things wrong. We can only hope to convince those who will listen to reasons, and those who do not care about others may not care about what others have to say, and hence may not listen to any reasons we give them. However, the purpose of moral philosophy is not to guarantee the assent of those who will not listen.

## 9.8 Cognitivism vindicated

Having explained how my account supports moral internalism, and having answered a number of objections, it is time to explain how the account supports moral cognitivism. It does so quite straightforwardly, because if we can act on moral reasons then we can have knowledge of moral reasons – the occasions on which we act on moral reasons are examples of cases in which we have such knowledge. Moreover, if we can judge what the moral reasons, taken together, require or favour, then we can make overall judgements about what is morally required. Even though the right-making relation is not the same as the requiring or favouring relation, both depend on moral reasons (see § 1.3).

It might be objected that I am confusing reasons for action and reasons for belief. Moral knowledge requires reasons for belief. Moral reasons are reasons for action. The reason why I ought to keep my promise is not necessarily the same as the reason why I should believe that I ought to keep my promise. This is true, but it does not show that the existence of moral reasons is insufficient to show that we can have moral knowledge. In many cases the considerations that give me reason to act also give me reason to have beliefs about what I ought to do. If you are in danger this can give me both a reason to believe that I ought to warn you and a reason to warn you.

At this point my account might seem to be vulnerable to Mackie's argument from queerness, since I claim that moral knowledge is possible, that moral demands are categorical, and that we must suppose that reasons and values are 'real', 'objective features of the world'. Mackie writes:

Plato's Forms give a dramatic picture of what objective values would have to be. The Form of the Good is such that knowledge of it provides the knower with both a direction and an overriding motive; something's being good both tells the person who knows this to pursue it and makes him pursue it. (1977 p. 40)

Korsgaard responds, writing:

Mackie is wrong and realism is right ... For it is the most familiar fact of human life that the world contains entities that can tell us what to do and make us do it. They are people, and the other animals (1996b p. 166).

I think that we should not accept Korsgaard's response; other people do not make us do things in the sense that Mackie intends. They do not provide us with an overriding motive that causes us to act, that guarantees our conformance. We can rebel against others, but on the Platonic picture that Mackie criticises our will is supposedly *determined* by our knowledge of the good.

Nonetheless, Mackie misrepresents the defender of moral objectivity, and does so by two conflation. Firstly, moral requirements might be reason giving without being overriding, but Mackie conflates these issues. Secondly, even if moral requirements were overriding, they would not necessarily make knowers conform to them by providing an overriding motive. Mackie conflates normativity and motivation: overriding requirements would fully determine what one *should* do, but would not determine one's actual actions and motivations. Overriding requirements are still ones that might go unheeded or be deliberately flouted. Moreover, moral requirements are not so ubiquitous and stringent that they determine what we must do in all situations. It is very rare that there is a single action that morality demands of us. We usually have a choice between a number of morally permissible alternatives. Holding that values are objective does not commit us to thinking that they *make us* pursue objects that instantiate them.

We can therefore conclude that the practical reason approach allows us to see that the existence of moral reasons entails that moral cognitivism is true. Moreover, since we have rejected the Humean theory of motivation, and found support for moral

internalism, we have not only defended both the objectivity of morality and the practicality of morality, but have also harmonised these elements.

## 9.9 Moral disagreement on the practical reason approach

A final challenge to moral cognitivism remains to be discussed: moral disagreement. The Kantian practical reason approach does not explain away moral disagreement, but it takes several important forms of moral disagreement to involve disagreement about reasons. We can disagree about whether a consideration is a contributory moral reason to  $\Phi$ , whether overall there is most moral reason to  $\Phi$ , about the relative weights of moral reasons, about whether a reason is moral or non-moral or about whether a non-moral reason may be taken into account. We can also disagree about whether other considerations defeat, enable, enhance or attenuate the normative force of a consideration. My suggestion is that moral disagreements can sometimes be solved by appeal to the moral law and the categorical imperative. If we are disputing whether a particular consideration justifies stealing, I can ask whether a maxim of stealing on the basis of that consideration is universalisable. Likewise, we can use the moral law to help resolve disputes. This will not help to resolve all our disagreements, however; it may not even help to resolve particularly many of them. If we are disagreeing about whether you should perform a particular action, or whether a given consideration supports acting in that way, we may find that, according to the categorical imperative and the moral law, it is permissible to perform that action for that reason, but also permissible not to. This may indicate that there is no fact of that matter about what should be done, but it does not necessarily do so since we are not supposing that all reasons arise from the categorical imperative and the moral law. The Kantian position does allow us guidance even here, however. According to the moral law we ought never to treat others as a mere means. This shows that when we disagree about what should be done, and the action is something that involves both of us, our decision needs to be taken together.

In discussing an anti-Humean or Kantian view that Smith takes to be the alternative to the Humeanism defended by Williams (1979), Smith writes:

[U]nder conditions of full rationality we would all reason ourselves towards the same conclusions as regards what is to be done... via a process of systematic justification of our desires we could bring it about that we converge in the desires that we have" (1994 pp. 165-6)

Whether it is true that full rationality would lead to convergence in desires or not, it is not a consequence of the Kantian view that this is so. Kantians might maintain that we are not required to give up our individuality; we are not all required to desire the same things. Kantians do not need to claim that there is a single best way of life, that all values are commensurable, and that we should all converge in our evaluations and desires. Rather, they claim that we need to respect each other's differences. When we deliberate about what to do, differences of many kinds (often grounded in our culture and upbringing) will make a great deal of difference to what we take into account. It is unsurprising therefore that we find that people have quite divergent ideas about what they should do and about what we should collectively do. It *may* be that everyone does have reason to do and want the same things, but this is something that needs to be argued for, not assumed, and it is no part of my project to argue for this claim in this thesis. Rather, my aim has to been to show that we should converge in our judgements about *morality* – we should all judge that we ought never to treat others as mere means, and that we should act only on maxims that can be willed as universal laws. These judgements make up a small subsection of our judgements about what we have reason to do, but a subsection that, if heeded, will ensure that we always treat others with the care and respect that they deserve. This is what morality requires of us. No matter what else we have reason to do, we have reason to do what morality requires.

## 9.10 Conclusions

I have argued that when we make moral judgements and act morally we recognise and respond to reasons that are there whether we recognise them or not. I have also claimed that morality requires that we treat others as ends in themselves rather than using them as mere means to achieve our ends. Whilst neither of these claims is particularly surprising from the point of view of commonsense morality, providing them with philosophical justification has proved notoriously difficult throughout the history of moral philosophy. This is, in part, due to a number of philosophical problems about the relationship between agency, motivation, rationality, knowledge and practical reason. By engaging with these problems we can gain a clearer understanding of the nature of morality and, if what I have said is right, possessing such understanding will help us to act morally by showing us how and when moral requirements give us reasons for action.

## Bibliography

- Abramson, Kate (2002) 'Two Portraits of the Humean Moral Agent' *Pacific Philosophical Quarterly* 83 pp. 301–334
- Agamben, Giorgio (1993) *The Coming Community* trans. Michael Hardt, Minneapolis: University of Minnesota Press
- Altham, J. E. J. (1986) 'The Legacy of Emotivism' in *Fact, Science, and Morality: Essays on A. J. Ayer's Language, Truth and Logic* (ed.) Graham Macdonald and Crispin Wright, Oxford: Blackwell
- Altham, J. E. J. and Harrison, Ross (eds.) (1995) *Mind, World and Ethics: Essays on the Ethical Philosophy of Bernard Williams* Cambridge: Cambridge University Press
- Anscombe, G. E. M. (1957) *Intention* Oxford: Blackwell
- Anscombe, G. E. M. (1958) Modern Moral Philosophy, in her *Ethics, Religion and Politics*, Oxford: Basil Blackwell 1981 pp. 26–42
- Appiah, Anthony (2006) *Cosmopolitanism: Ethics in a World of Strangers* New York: W.W. Norton & Co.
- Aristotle (1998) *Nicomachean Ethics* Ontario: Dover Thrift Editions
- Armstrong, David M., Martin, Charlie B., Place, U. T. (1996) *Dispositions: A Debate*, London: Routledge
- Audi, Robert (1990) 'Rationality and Valuation' in Paul K. Moser (ed.) (1990) *Rationality in Action: Contemporary Approaches* Cambridge: Cambridge University Press, pp. 416–446
- Audi, Robert (1996) 'Objectivity and the Internal-External Reasons Controversy: A Study of Paul K. Moser's *Philosophy after Objectivity*' *Philosophical and Phenomenological Research* LVI (2) pp. 395–400
- Audi, Robert (1997) 'Moral Judgement and Reasons for Action' in Garrett Cullity and Berys Gaut (eds.) *Ethics and Practical Reason* Oxford: Oxford University Press, pp. 125–159
- Ayer, A. J. (1946) *Language, Truth, and Logic*, 2<sup>nd</sup> Edition, London: Gollancz, first published 1936
- Ayer, A. J. (1954) 'Freedom and Necessity' reprinted in Gary Watson (ed.) 1982 *Free Will* Oxford: Oxford University Press, pp. 15–23
- Baier, Annette (1989) 'Doing without Moral Theory?' in Stanley G. Clarke and Evan Simpson (eds.) (1989) *Anti-Theory and Moral Conservatism in Ethics*, Albany: State University of New York Press
- Baier, Annette (1991) *A Progress of Sentiments: Reflections on Hume's Treatise* Cambridge: Harvard University Press
- Barnes, W. H. F., (1933) 'A Suggestion About Value,' *Analysis*, 1 pp. 45–46
- Barry, Melissa (2007) 'Realism, Rational Action, and the Humean Theory of Motivation' *Ethical Theory and Moral Practice* 10 pp. 231–242
- Bedke, M. S. (2009) 'Moral Judgment Purposivism: Saving Internalism from Amoralism' *Philosophical Studies* 144 pp. 189–209
- Bennett, Jonathan (1988) *Events and Their Names* Indianapolis: Hackett
- Bittner, Rüdiger (2001) *Doing Things for Reasons* Oxford: Oxford University Press
- Bird, Alexander (1998) 'Dispositions and Antidotes' *The Philosophical Quarterly* 48 pp. 227–234
- Blackburn, Simon (1984) *Spreading the Word* Oxford: Oxford University Press

- Blackburn, Simon (1985) 'Errors and the Phenomenology of Value', in Honderich 1985, reprinted in Blackburn 1993a
- Blackburn, Simon (1988) 'Attitudes and Contents', *Ethics*, 98 pp. 501–17
- Blackburn, Simon (1992) 'Through Thick and Thin' *Proceedings of the Aristotelian Society*, Supplementary Volume 66 pp. 285–99.
- Blackburn, Simon (1993a) *Essays in Quasi-Realism*, New York: Oxford University Press
- Blackburn, Simon (1993b) 'Realism, Quasi or Queasy?', in John Haldane and Crispin Wright (eds.) (1993) *Reality, Representation and Projection* Oxford: Oxford University Press
- Blackburn, Simon (1996) 'Securing the Nots', in Walter Sinnott-Armstrong and Mark Timmons (eds.) (1996) *Moral Knowledge?* Oxford: Oxford University Press
- Blackburn, Simon (1998) *Ruling Passions* Oxford: Oxford University Press
- Bracken, Patrick J. (1999) 'The Importance of Heidegger for Psychiatry' *Philosophy, Psychiatry and Psychology* 6 pp. 83–85
- Brady, Michael S. (2000) 'How to Understand Internalism' *The Philosophical Quarterly*, 50 pp. 91–97
- Brandt, Richard B. (1990) 'The Concept of Rational Action' in Paul K. Moser (ed.) (1990) *Rationality in Action: Contemporary Approaches* Cambridge: Cambridge University Press, pp. 398–415
- Bratman, Michael (1983) 'Taking Plans Seriously', *Social Theory and Practice* 9 (2–3) pp. 271–287
- Bratman, Michael E. (1999) *Faces of Intention* Cambridge: Cambridge University Press
- Bricke, John (1996) *Mind and Morality: An Examination of Hume's Moral Psychology* Oxford: Oxford University Press
- Brink, David O. (1997) 'Moral Motivation' *Ethics* 108 pp. 4–32
- Britton, Karl (1939) *Communication: A Philosophical Study of Language* York: Harcourt Brace
- Broad (1933–34) 'Is Goodness the Name of a Simple, Non-Natural Property?' *Proceedings of the Aristotelian Society* 34 pp. 249–68
- Broome, John (2000) 'Normative Requirements', in Jonathan Dancy (ed.) (2000b) *Normativity* Oxford: Blackwell
- Broome, John (2004) 'Reasons' in Wallace, R. Jay, Pettit, Philip, Scheffler, S., Smith, Michael (eds.) *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* Oxford: Clarendon Press
- Broome, John (2007a) 'Does Rationality Consist in Responding Correctly to Reasons?' *Journal of Moral Philosophy* 4 pp. 349–74
- Broome, John (2007b) 'Is Rationality Normative?' *Disputatio*, II (23) (special issue) pp. 161–78
- Butler, Joseph (1914) *Fifteen Sermons Preached at the Rolls Chapel and A Dissertation Upon the Nature of Virtue* London: G. Bell and Sons Ltd
- Carnap, Rudolph (1935) *Philosophy and Logical Syntax* London: Kegan Paul
- Churchland, Paul (1979) *Scientific Realism and the Plasticity of Mind*, Cambridge: Cambridge University Press
- Churchland, Paul (1981) 'Eliminative Materialism and the Propositional Attitudes' *Journal of Philosophy* 78 pp. 67–90
- Clarke, Stanley G. and Simpson, Evan (eds.) (1989) *Anti-Theory and Moral Conservatism in Ethics* Albany: State University of New York Press
- Cohen, Gerald A. (1996) 'Reason, Humanity, and the Moral Law' in Onora O'Neill (ed.) *The Sources of Normativity* Cambridge: Cambridge University Press

- Cohen, Gerald A. (2003) 'Facts and Principles', *Philosophy and Public Affairs* 31 pp. 211-45.
- Cohon, Rachel (1986) 'Are External Reasons Impossible?' *Ethics* 96 pp. 545-556
- Cooper, David (1990) *Existentialism: A Reconstruction* Basil Blackwell: Oxford
- Cooper, David (ed.) (1998) *Ethics: The Classic Readings* Blackwell: Oxford
- Copp, David (1991) 'Moral Skepticism', *Philosophical Studies*, 62 pp. 203-33
- Copp, David (1995) *Morality, Normativity, and Society* Oxford: Oxford University Press
- Copp, David (1997) 'Belief, Reason, and Motivation: Michael Smith's "The Moral Problem"', *Ethics*, 108, pp. 33-54
- Copp, David (ed.) (2006) *The Oxford Handbook of Ethical Theory* Oxford: Oxford University Press
- Cowley, Christopher (2005) 'A New Defence of Williams' Reasons-Internalism' *Philosophical Investigations* 28 (4) pp. 346-368
- Crary, Alice (2002) 'Why Can't Moral Thought Be Everything It Seems?' *The Philosophical Forum* 33 pp. 373-391
- Crisp, Roger (2000) 'Particularizing Particularism', in Hooker and Little (eds.) 2000 *Moral Particularism* Oxford: Oxford University Press, pp. 23-47
- Cullity, Garrett and Gaut, Berys (eds.) (1997) *Ethics and Practical Reason*, Oxford: Oxford University Press
- Curry, Oliver (2005) *Morality as Natural History*, PhD thesis, London: LSE Research Online. Available at: <http://eprints.lse.ac.uk/archive/00000441>
- Dahl, Norman O. (2009) 'Aristotle on Action, Practical Reason, and Weakness of the Will' in Georgios Anagnostopoulos (ed.) *A Companion to Aristotle*, Oxford: Wiley-Blackwell, pp. 498-511
- Dancy, Jonathan (1993) *Moral Reasons* Oxford: Blackwell
- Dancy, Jonathan (1995) 'In Defence of Thick Concepts' *Midwest Studies in Philosophy* 20 pp. 263-79
- Dancy, Jonathan (2000a) *Practical Reality* Oxford: Oxford University Press
- Dancy, Jonathan (ed.) (2000b) *Normativity* Oxford: Blackwell
- Dancy, Jonathan (2003a) 'Precis of *Practical Reality*' *Philosophy and Phenomenological Research* LXVII pp. 423-28
- Dancy, Jonathan (2003b) 'Replies' *Philosophy and Phenomenological Research* LXVII pp. 468-90
- Dancy, Jonathan (2004) *Ethics Without Principles* Oxford: Oxford University Press
- Dancy, Jonathan (2005) 'Moral Particularism' in Edward N. Zalta (ed.), *The Stanford Encyclopaedia of Philosophy* (Summer 2005 Edition), <http://plato.stanford.edu/archives/sum2005/entries/moral-particularism/>
- Dancy, Jonathan (2007) 'Defending the Right' *Journal of Moral Philosophy* 4 pp. 85-98
- Darwall, Stephen (1983) *Impartial Reason* Ithaca: Cornell University Press
- Darwall, Stephen, Gibbard, Allan and Railton, Peter (1992) 'Towards *Fin de siècle* Ethics: Some Trends', *Philosophical Review* pp. 115-189
- Darwall, Stephen (2006) *Morality and Practical Reasons: A Kantian Approach* in David Copp (ed.) 2006 *The Oxford Handbook of Ethical Theory* Oxford: Oxford University Press
- Davies, M. and Stone T. (eds.) (1995) *Folk Psychology: The Theory of Mind Debate*, Oxford: Blackwell
- Davidson, Donald (1963) 'Actions, Reasons and Causes' *The Journal of Philosophy* LX pp. 685-700



- Davidson, Donald (1970) 'Mental Events' in Lawrence Foster and J. W. Swanson (eds.) *Experience and Theory* London: Duckworth
- Davidson, Donald (1978) 'Intending' reprinted in Davidson 1980 pp. 83-102
- Davidson, Donald (1980) *Essays on Actions and Events*, Oxford: Oxford University Press
- Davidson, Donald (1984) *Inquiries into Truth and Interpretation* Oxford: Oxford University Press
- Davidson, Donald (2004) 'The Objectivity of Values' in his *Problems of Rationality*, Oxford: Clarendon Press
- Dennett, Daniel (1978) *Brainstorms* Montgomery, VT: Bradford Press
- Dennett, Daniel (1987) *The Intentional Stance* Cambridge, MA: MIT Press
- Dewey, John (1922) *Human Nature and Conduct: An Introduction to Social Psychology* New York: Modern Library
- Donagan, Alan (1987) *Choice: The Essential Element in Human Action* London: Routledge & Kegan Paul
- Dreier, James (1997) 'Humean Doubts about the Practical Justification of Morality' in Garrett Cullity and Berys Gaut (eds.) *Ethics and Practical Reason*, Oxford: Oxford University Press, pp. 81-99
- Dreier, James (1999) 'Transforming Expressivism' *Noûs* 33 pp. 558–572
- Dreier, James (2000) 'Dispositions and Fetishes: Externalist Models of Moral Motivation' *Philosophy and Phenomenological Research* 61 pp. 619-38
- Dreier, James (2002) 'Troubling Developments in Metaethics' *Noûs* 36 pp. 152–168
- Dreier, James (ed.) (2006) *Contemporary Debates in Moral Theory* Oxford: Blackwell
- Dworkin, Ronald (1996) 'Objectivity and Truth: You'd Better Believe It,' *Philosophy and Public Affairs* 25, pp. 87-139
- Elstein, Daniel (2006) 'The Open Question Argument and Cornell Realism' presented at *Problems and Prospects for Ethical Naturalism*, Durham University, 11th – 12th August 2006
- Elstein, Daniel (2008) 'Is There a Normative Question, and If So How Can It Be Answered?' Oxford Moral Philosophy Seminar 5<sup>th</sup> May 2008, available from <http://ethics-etc.com/wp-content/uploads/2008/05/elstein.pdf> under the title 'Humean and Kantian Approaches to Normative Justification', retrieved 12/04/10
- Enoch, David (2006) 'Agency, Schmagency: Why Normativity Won't Come from What Is Constitutive of Action' *Philosophical Review* 115 pp. 169-98
- Enoch, David (2009) 'How is Moral Disagreement a Problem for Realism?' *Journal of Ethics* 13 pp. 15–50
- Eriksson, John (2009) 'Homage to Hare: Ecumenism and the Frege-Geach Problem' *Ethics* 120 pp. 8-35
- Fantl, Jeremy (2006) 'Is Metaethics Morally Neutral?' *Pacific Philosophical Quarterly* 87 pp. 24–44
- Fantl, Jeremy (2008) 'Knowing-How and Knowing-That' *Philosophy Compass* 3 pp. 451–70
- Fara, Michael (2005) 'Dispositions and Habituals', *Noûs* 39 pp. 43–82
- Fara, Michael (2009) 'Dispositions', The Stanford Encyclopedia of Philosophy (Summer 2009 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2009/entries/dispositions/>>.
- Fehige, Christoph (2001) 'Instrumentalism' in Elijah Millgram (ed.) 2001 *Varieties of Practical Reasoning*, Cambridge MA: MIT Press

- FitzPatrick, William J. (2005) 'The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity', *Ethics* 115 pp. 651–691
- Foot, Philippa. (1972) 'Morality as a System of Hypothetical Imperatives' *The Philosophical Review*, 81 pp. 305–316
- Foot, Philippa (1985) 'Utilitarianism and the Virtues' *Mind* 94 pp. 196–209
- Frankena, Williams K. (1938) 'The Naturalistic Fallacy' *Mind* 48 pp. 464–77
- Frankena, William K. (1958) 'Obligation and Motivation in Recent Moral Philosophy' in A. I. Melden, (ed.) 1958, *Essays on Moral Philosophy* Seattle: University of Washington Press, pp. 40–81
- Frankfurt, Harry (1971) 'Freedom of the Will and the Concept of a Person' reprinted in Gary Watson (ed.) 1982, *Free Will*, Oxford: Oxford University Press pp. 81–95
- Frankfurt, Harry (1988) *The Importance of What We Care About* Cambridge: Cambridge University Press
- Frankfurt, Harry (1993) 'On the Necessity of Ideals' in G. Noam & T. Wren (eds.), *The Moral Self*, Cambridge: MIT Press, pp. 108–16, reprinted in Harry Frankfurt (1999) *Necessity, Volition, and Love* Cambridge: Cambridge University Press
- Frankfurt, Harry (1999) 'On Caring' in his *Necessity, Volition, and Love* Cambridge: Cambridge University Press
- Frege, Gottlob (1892) 'Über Sinn und Bedeutung', in *Zeitschrift für Philosophie und philosophische Kritik*, 100 pp. 25–50, translated as 'On Sense and Reference' by Max Black in *Translations from the Philosophical Writings of Gottlob Frege*, Peter Geach and Max Black (eds. and trans.), Oxford: Blackwell, third edition, 1980
- Freud, Sigmund (1914) 'Psychopathology of Everyday Life', trans. A. A. Brill, London: Ernest Benn
- Freud, Sigmund (1960) *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, trans. James Strachey et al. London: Hogarth Press
- Frith, Uta and Happe, Francesca (1999) 'Theory of Mind and Self-Consciousness: What Is It Like to Be Autistic?' *Mind & Language* 14 pp 1–22
- Gardner, Sebastian, (1993) *Irrationality and the Philosophy of Psychoanalysis* Cambridge: Cambridge University Press
- Garrard, Eve and McNaughton, David (1998) 'Mapping Moral Motivation' *Ethical Theory and Moral Practice*, 1, pp. 45–59
- Geach, Peter T. (1956) 'Good and Evil' *Analysis* 17 pp. 32–42
- Geach, Peter T. (1958) 'Imperative and Deontic Logic' *Analysis* 18 pp. 49–56
- Geach, Peter T. (1960) 'Ascriptivism' *Philosophical Review* 69 pp. 221–5
- Geach, Peter T. (1965) 'Assertion' *Philosophical Review* 74 pp. 449–65
- Geertz, Clifford (1973) 'Thick Description: Towards an Interpretive Theory of Culture' in his *Interpretation of Cultures* New York: Basic Books
- Gert, Joshua (2002) 'Avoiding the Conditional Fallacy' *The Philosophical Quarterly* 52 pp. 88–95
- Gibbard, Allan (1990) *Wise Choices, Apt Feelings* Cambridge, MA: Harvard University Press
- Gibbard, Allan (1992) 'Thick Concepts and Warrant for Feelings' *Proceedings of the Aristotelian Society*, Supplementary Volume 66 pp. 267–83
- Gibbard, Allan (2006) 'Normative Properties' in Terry Horgan and Mark Timmons (eds.) *Metaethics After Moore* Oxford: Oxford University Press
- Ginet, Carl (1975) *Knowledge, Perception, and Memory* Dordrecht: D. Reidel Publishing Company

- Goldie, Peter (2007) 'There are Reasons and Reasons' in Matthew Ratcliffe and Daniel Hutto (eds.) (2007) *Folk Psychology Reassessed* Dordrecht: Springer
- Goldman, Alvin I. (1970) *A Theory of Human Action* Englewood Cliffs, New Jersey: Prentice-Hall
- Goldstein, Peter (2004) 'An Argument Against Williams' Internalism', *Logos* 2 pp.60-81
- Gowans, Christopher W. (ed.) (2000) *Moral Disagreements: Classic and Contemporary Readings* London: Routledge
- Gray, John (1995) *Isaiah Berlin* London: Fontana
- Gray, John (2000) *Two Faces of Liberalism* Cambridge: Polity
- Guignon, Charles B. (1993) (ed.) *The Cambridge Companion to Heidegger* Cambridge: Cambridge University Press
- Haldane, John and Wright, Crispin (eds.) (1993) *Reality, Representation and Projection* Oxford: Oxford University Press
- Hale, Bob (1986) 'The Compleat Projectivist' *Philosophical Quarterly* 36 pp. 65-84
- Hale, Bob (1993) 'Can There be a Logic of Attitudes?' in John Haldane and Crispin Wright (eds.) (1993) *Reality, Representation and Projection* Oxford: Oxford University Press
- Hall, Harrison (1993) 'Intentionality and world: Division I of *Being and Time*' in Charles B. Guignon (ed.) *The Cambridge Companion to Heidegger* Cambridge: Cambridge University Press, pp. 122-140
- Harding, Sarah (1999) 'Value, Obligation and Cultural Heritage', *Arizona State Law Journal* 31 pp. 291-354
- Hare, R. M. (1952) *The Language of Morals* Oxford: Oxford University Press
- Hare, R. M. (1981) *Moral Thinking: Its Levels, Methods, and Point* Oxford: Oxford University Press
- Harman, Gilbert (1977) *The Nature of Morality* Oxford: Oxford University Press.
- Harrison, Ross (ed.) (1979) *Rational Action: Studies in Philosophy and Social Science* Cambridge: Cambridge University Press.
- Hartland-Swann, John (1956) 'The Logical Status of "Knowing That" ' *Analysis* 16 (5) pp. 111-15
- Heidegger, Martin (1962) *Being and Time* trans. John Macquarrie and Edward Robinson, New York: Harper & Row
- Hempel, Carl G. (1961) 'Rational Action' *Proceedings and Addresses of the American Philosophical Association* 35 pp. 5-23
- Hetherington, Stephen (2006) 'How to Know (that Knowledge-That is Knowledge-How)' in Stephen Hetherington (ed.) *Epistemology Futures* Oxford: Clarendon Press, pp. 71-94
- Hobbes, Thomas (1651) *Leviathan*, in E. Curley (ed.), *Leviathan, with selected variants from the Latin edition of 1668*, Indianapolis: Hackett, 1994
- Honderich, Ted (ed.) (1985) *Morality and Objectivity: A Tribute to John Mackie* London: Routledge & Kegan Paul
- Hooker, Brad (1987) 'Williams' Argument against External Reasons', *Analysis*, 47 (1) pp. 42-44
- Hooker, Brad and Little, Margaret Olivia (eds.) (2000) *Moral Particularism* Oxford: Oxford University Press
- Horgan, Terry and Timmons, Mark (1992) 'Troubles for New Wave Moral Semantics: the "Open-Question Argument" Revived' *Philosophical Papers* 21 pp. 153-75
- Horgan, Terry and Timmons, Mark (eds.) (2006) *Metaethics after Moore* Oxford: Oxford University Press

- Hornsby, Jennifer (1980) *Actions* London: Routledge and Kegan Paul
- Hudson, W. D. (1969) (ed.) *The Is/Ought Question* London: Macmillan
- Hume, David (1739-1740) *A Treatise of Human Nature*, (ed.) L. A. Selby-Bigge and P. H. Nidditch (2<sup>nd</sup> edition) Oxford: Oxford University Press, 1978
- Huoranszki, Ferenc (2002) 'Common Sense and the Theory of Human Behaviour' *The Philosophical Quarterly*, 52 pp. 526-43
- Hutcheson, Francis (1728) *Illustrations upon the Moral Sense*, sect. I; repr. abridged in L. A. Selby-Bigge (ed.) 1897 *British Moralists*, 2 vols. Oxford: Clarendon Press, i. 403-418
- Hutto, Daniel D. (2007) 'Folk Psychology without Theory or Simulation' in Matthew Ratcliffe and Daniel Hutto (eds.) (2007) *Folk Psychology Reassessed* Dordrecht: Springer
- Hutto, Daniel D and Ratcliffe, Matthew (eds.) (2007) *Folk Psychology Reassessed* Dordrecht: Springer
- Izzard, Eddie (1998) *Dress To Kill* Ella Communications
- Jackson, Frank (1992) 'Block's Challenge' in K Campbell, J Bacon, L Rhinehart (eds.) *Ontology, Causality, and Mind: Essays on the Philosophy of David Armstrong* Cambridge: Cambridge University Press
- James, Henry (1992) 'Owen Wingrave' in Henry James 1992 *The Turn of the Screw and Other Stories*, Oxford: Oxford University Press
- Johnson, Robert N. (1999) 'Internal Reasons and the Conditional Fallacy' *The Philosophical Quarterly* 49 (194) pp. 53-71
- Johnson, Robert N. (2003) 'Internal Reasons: Reply to Brady, van Roojen and Gert' *The Philosophical Quarterly* 53 (213) pp. 573-580
- Johnston, Mark (1992) 'How to Speak of the Colors', *Philosophical Studies* 68 pp. 221-263
- Jones, Peter (2006) 'Toleration, Value Pluralism, and the Fact of Pluralism', *Critical Review of International Social and Political Philosophy* 9 (2) pp.189-210
- Kant, Immanuel (1785) *Grounding for the Metaphysics of Morals* trans. James W. Ellington, 3<sup>rd</sup> edition, 1981, Indiana: Hackett
- Kant, Immanuel (1909) *Critique of Practical Reason* trans. Thomas Kingsmill Abbott, eds. Rosenkranz and Schubert, London: Longmans, Green
- Kaplan, Abraham (1942) 'Are Moral Judgments Assertions?' *Philosophical Review* 51 pp. 280-303
- Keynes, John Maynard (1972) 'My Early Beliefs' in *The Collected Works of John Maynard Keynes*, vol x., Cambridge: Macmillan
- Kirchin, Simon (2007) 'Particularism and Default Valency' *Journal of Moral Philosophy* 4 pp. 16-32
- Knobe, Joshua (2007) 'Folk Psychology: Science and Morals' in Matthew Ratcliffe and Daniel Hutto (eds.) (2007) *Folk Psychology Reassessed* Dordrecht: Springer
- Koethe, John (2002) 'Stanley and Williamson on Knowing How' *The Journal of Philosophy* 99 (6) pp. 325-8
- Korsgaard, Christine (1986) 'Skepticism about Practical Reason', *Journal of Philosophy* 83 pp. 5-25, reprinted in Korsgaard (1996a)
- Korsgaard, Christine (1996a) *Creating the Kingdom of Ends* Cambridge: Cambridge University Press
- Korsgaard, Christine (1996b) *The Sources of Normativity* Cambridge: Cambridge University Press

- Korsgaard, Christine (2003) 'Realism and Constructivism in Twentieth-Century Moral Philosophy', APA Centennial Supplement to the *Journal of Philosophical Research* pp. 99-122
- Korsgaard, Christine (2008) *The Constitution of Agency: Essays on Practical Reason and Moral Psychology* Oxford: Oxford University Press
- Korsgaard, Christine (2009) *Self-Constitution: Agency, Identity and Integrity* Oxford: Oxford University Press
- Kripke, Saul (1980) *Naming and Necessity*, 2<sup>nd</sup> edition, Cambridge MA: Harvard University Press
- Kripke, Saul (1982) *Wittgenstein on Rules and Private Language* Cambridge, Harvard University Press
- Kuhse, Hegla (1991) 'Euthanasia' in Peter Singer (ed.) *A Companion to Ethics* Oxford: Blackwell
- Kusch, Martin (2007) 'Folk Psychology and Freedom of the Will' in Matthew Ratcliffe and Daniel Hutto (eds.) (2007) *Folk Psychology Reassessed* Dordrecht: Springer
- Laidlaw, James (2002) 'For an Anthropology of Ethics and Freedom', *Journal of the Royal Anthropological Institute* 8 (2) pp. 311-32.
- Lenman, James (1996) 'Belief, Desire and Motivation: An Essay in Quasi-Hydraulics' *American Philosophical Quarterly* 33 pp. 291-301
- Lewis, David (1997) 'Finkish Dispositions' *The Philosophical Quarterly* 47 pp. 143-58
- Lillard, Angeline (1997) 'Other Folks' Theories of Mind and Behavior' *Psychological Science*, 8 pp. 268-274
- Lillard, Angeline (1998) 'Ethnopsychologies: Cultural Variations in Theories of Mind' *Psychological Bulletin* 123 pp. 3-32
- Lillehammer, Hallvard (2000) 'The Doctrine of Internal Reasons' *The Journal of Value Inquiry* (394) pp. 507-16
- Little, Margaret Olivia (1997) 'Virtue as Knowledge: Objections from the Philosophy of Mind' *Noûs* 31 pp. 59-79.
- Little, Margaret Olivia (2001) 'On Knowing the 'Why': Particularism and Moral Theory' *Hasting Centre Report* 31 (4) pp.32-40.
- Lockie, Robert (1998) 'What's Wrong with Moral Internalism' *Ratio (new series)* XI pp. 14-36
- Lovibond, Sabina (2002) *Ethical Formation* Cambridge, MA: Harvard University Press
- Lowe, Edward Jonathan (2008) *Personal Agency: The Metaphysics of Mind and Action* Oxford: Oxford University Press
- Mackie, John L. (1977) *Ethics: Inventing Right and Wrong* London: Penguin Books
- MacIntyre, Alasdair (1967) *A Short History of Ethics*, 2<sup>nd</sup> Edition (1998) London: Routledge
- MacIntyre, Alasdair (1981) *After Virtue: A Study in Moral Theory* London: Duckworth
- Mahmood, Saba (2005) *Politics of Piety: The Islamic Revival and the Feminist Subject* Princeton, NJ: Princeton University Press.
- Martin, Charlie B. (1994) 'Dispositions and Conditionals' *The Philosophical Quarterly* 44 (174) pp. 1-8
- Mason, Elinor (2008) 'An Argument Against Motivational Internalism' *Proceedings of the Aristotelian Society*, Vol. cviii, Part 2, pp. 135-156
- McDowell, John (1978) 'Are Moral Requirements Categorical Imperatives?' *Proceedings of the Aristotelian Society Supplementary Volume* pp. 13-29
- McDowell, John (1979) 'Virtue and Reason', *The Monist*, pp. 331-50

- McDowell, John (1981) 'Non-Cognitivism and Rule Following' in Steven Holtzman and Christopher Leich (eds.) *Wittgenstein: To Follow a Rule* London: Routledge pp. 141-62
- McDowell, John (1994) *Mind and World* Cambridge, MA: Harvard University Press
- McDowell, John (1995) 'Might There Be External Reasons?' in J. E. J. Altham and Ross Harrison (eds.) 1995 *Mind, World and Ethics: Essays on the Ethical Philosophy of Bernard Williams* Cambridge: Cambridge University Press, pp. 68-85
- McDowell, John (1998) *Mind, Value and Reality* Cambridge, MA: Harvard University Press
- McGeer, Victoria (2007) 'The Regulative Dimension of Folk Psychology' in Dan Hutto and Matthew Ratcliffe (eds.) (2007) *Folk Psychology Re-assessed*, Dordrecht: Springer, pp. 137-156
- McKeever, Sean and Ridge, Michael (2006) *Principled Ethics* Oxford: Oxford University Press
- McLaughlin, Brian and Bennett, Karen (2010) 'Supervenience' *The Stanford Encyclopedia of Philosophy* (Summer 2010 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2010/entries/supervenience/>>.
- McNaughton, David (1988) *Moral Vision* Oxford: Basil Blackwell
- Melden, A. I. (1961) *Free Action* London: Routledge and Kegan Paul
- Mele, Alfred R. (1992) *Springs of Action* Oxford: Oxford University Press
- Mele, Alfred (1995) 'Motivation: Essentially Motivation-Constituting Attitudes' *The Philosophical Review*, 104, pp. 387-423
- Mehl, Peter J. (1986) 'Kierkegaard and the Relativist Challenge to Practical Philosophy' *The Journal of Religious Ethics*, 14 (2) pp. 247-278
- Miller, Alexander (2003) *An Introduction to Contemporary Metaethics*, Cambridge: Polity
- Miller, Christian Basil (2008) 'Motivational Internalism' *Philosophical Studies* 139 pp. 233-255
- Millgram, Elijah (1995) 'Was Hume a Humean?' *Hume Studies*, XXI, pp. 75-94
- Millgram, Elijah (ed.) (2001) *Varieties of Practical Reasoning*, Cambridge MA: MIT Press
- Molnar, George (2003) *Powers: A Study in Metaphysics* Oxford: Oxford University Press
- Moore, George Edward (1903) *Principia Ethica* London: Cambridge University Press
- Morton, Adam (1970) *Frames of Mind* Oxford: Oxford University Press
- Morton, Adam (2003) *The Importance of Being Understood: Folk Psychology as Ethics* London: Routledge
- Morton, Adam (2007) 'Folk Psychology Does Not Exist' in Matthew Ratcliffe and Daniel Hutto (eds.) (2007) *Folk Psychology Reassessed* Dordrecht: Springer
- Moser, Paul K. (1993) *Philosophy after Objectivity* New York and Oxford: Oxford University Press
- Mumford, Stephen (1998) *Dispositions* Oxford: Oxford University Press.
- Nagel, Thomas (1970) *The Possibility of Altruism* Princeton, N. J.: Princeton University Press
- Nagel, Thomas (1986) *The View from Nowhere* Oxford: Oxford University Press.
- Nagel, Thomas (1997) *The Last Word* Oxford: Oxford University Press
- Needham, Rodney (1972) *Belief, Language and Experience* Oxford: Blackwell
- Nietzsche, Friedrich (1887) *The Gay Science* trans. Walter Kaufmann, 1974, New York: Vintage Books

- Nowell-Smith, Patrick Horace (1954) *Ethics* Harmondsworth: Pelican, Penguin Books
- Nussbaum, Martha (2000) 'Why Practice Needs Ethical Theory: Particularism, Principle and Bad Behaviour' in Brad Hooker and Olivia Margaret Little (eds.) 2000 *Moral Particularism* Oxford: Oxford University Press
- Ogden, C. K. and Richards, I. A., (1923) *The Meaning of Meaning*, New York: Harcourt Brace & Jovanovich.
- O'Shaughnessy, Brian (1980) *The Will*, vol. ii, Cambridge: Cambridge University Press
- Owen, David (2000) *Hume's Reason* Oxford: Oxford University Press
- Parekh, Bhikhu (2000) *Rethinking Multiculturalism: Cultural Diversity and Political Theory* Cambridge, MA: Harvard University Press
- Parfit, Derek (1984) *Reasons and Persons*, Oxford: Oxford University Press
- Parfit, Derek (1997) 'Reasons and Motivation', *The Aristotelian Society*, Supplementary Volumes 77 pp. 99-130
- Perry, Charner M. (1933) 'Bases, Arbitrary and Otherwise, for Morality: A Critique Criticized: The Arbitrary as Basis for Rational Morality' *International Journal of Ethics* 63 pp. 127-44
- Peters, R. S. (1958) *The Concept of Motivation*. London: Routledge and Kegan Paul
- Pettit, Philip (1987) 'Humeans, Anti-Humeans, and Motivation', *Mind* 96 (384) pp. 530-3
- Pettit, Philip and Smith, Michael (1990) 'Backgrounding Desire' *The Philosophical Review*, XCIX pp. 565-592
- Piller, Christian (2003) 'Two Accounts of Objective Reasons' *Philosophy and Phenomenological Research* 67 pp. 444-451
- Piller, Christian (2006) 'Particularism and the Structure of Reasons' *Acta Analytica* 21 pp. 87-102
- Pincoffs, Edmund (1971) 'Quandry Ethics', *Mind* 80 (320) pp.552-571
- Plato (1993) *Republic* trans. Robin Waterfield, Oxford: Oxford University Press
- Platts, Mark (1979) *Ways of Meaning* London: Routledge and Kegan Paul
- Platts, Mark (1981) 'Moral Reality and the End of Desire' in Mark Platts, ed., *Reference, Truth and Reality*. Routledge and Kegan Paul pp. 69-82
- Platts, Mark (1988) 'Hume and Morality as a Matter of Fact' *Mind*, 97, pp. 189-204.
- Pollard, Bill (2006) 'Actions, Habits and Constitution', *Ratio* XIX (2) pp. 229-248
- Price, Richard (1787) *A Review of the Principle Questions in Morals* 3<sup>rd</sup> Edition, London: T. Cadell
- Prichard, Harold Arthur (1912) 'Does Moral Philosophy Rest on a Mistake?' *Mind* 21 pp. 21-37
- Prichard, Harold Arthur (1945) 'Acting, Willing, Desiring' in his *Moral Writings*, Jim MacAdam (ed.) 2002, Oxford: Clarendon Press, pp. 272-81
- Putnam, Hilary (1975) 'The Meaning of "Meaning"', reprinted in his *Mind, Language and Reality* Cambridge: Cambridge University Press, pp. 215-71.
- Rachels, James (ed.) (1998) *Ethical Theory* Oxford: Oxford University Press
- Radcliffe, Elizabeth S. (2008) 'The Humean Theory of Motivation and its Critics' in *A Companion to Hume* (ed.) Elizabeth S. Radcliffe, Oxford: Blackwell Publishing, 2008 Blackwell Reference Online 18 December 2009 [http://www.blackwellreference.com/subscriber/tocnode?id=g9781405114554\\_c\\_hunk\\_g978140511455429](http://www.blackwellreference.com/subscriber/tocnode?id=g9781405114554_c_hunk_g978140511455429)
- Radzik, L. (2000) 'Justification and the Authority of Norms', *Journal of Value Inquiry* 34(4) pp. 451-61.
- Railton, Peter (1986) 'Facts and Values' *Philosophical Topics* 14 pp. 15-31

- Railton, Peter (2000) 'Normative Force and Normative Freedom', in Jonathan Dancy (ed.) (2000b) *Normativity* Oxford: Blackwell
- Ratcliffe, Matthew (2007a) *Rethinking Commonsense Psychology: A Critique of Folk Psychology, Theory of Mind, and Simulation* Basingstoke: Palgrave Macmillan
- Ratcliffe, Matthew (2007b) 'From Folk Psychology to Commonsense', in Matthew Ratcliffe and Daniel D. Hutto (eds.) (2007) *Folk Psychology Reassessed* Dordrecht: Springer
- Ravenscroft, Ian (2003) 'Simulation, Collapse and Humean Motivation' *Mind & Language* 18 pp. 162-174
- Raz, Joseph (ed.) (1978) *Practical Reason* Oxford: Oxford University Press
- Raz, Joseph (1997) 'The Amoralist' in Garrett Cullity and Berys Gaut (eds.) *Ethics and Practical Reason*, Oxford: Oxford University Press, pp. 369-398
- Raz, Joseph (2000) 'Explaining Normativity: On Rationality and the Justification of Reason', in Jonathan Dancy (ed.) (2000b) *Normativity* Oxford: Blackwell
- Rhees, Rush (1969) *Without Answers* London: Routledge and Kegan Paul
- Ridge, Michael (2006) 'Ecumenical Expressivism: Finessing Frege' *Ethics* 116 pp. 302-336
- Robbins, Joel (2004) *Becoming Sinners: Christianity and Moral Torment in a Papua New Guinea Society* Berkeley: University of California Press.
- Roland, Jane (1958) 'On "Knowing How" and "Knowing That" ' *The Philosophical Review* 67 (3) pp. 379-88.
- Rorty, Richard, (1979) *Philosophy and the Mirror of Nature* Princeton, N. J.: Princeton University Press
- Rorty, Richard, (1989) *Contingency Irony and Solidarity* Cambridge: Cambridge University Press
- Rosati, Connie (2003) 'Agency and the Open Question Argument' *Ethics* 113 pp. 490-527
- Ross, William David (1930) *The Right and the Good* Oxford: Clarendon Press
- Rousseau, Jean-Jacques (1968) *The Social Contract* trans. Maurice Cranston, Baltimore: Penguin
- Ruben, David-Hillel (1997) 'The Active and the Passive II' *Proceedings of the Aristotelian Society*, Supplementary Volumes, Vol. 71, pp. 229-246
- Russell, Bertrand (1935) *Religion and Science* New York: Henry Holt
- Russell, Paul (2004) 'Practical Reason and Motivational Scepticism' *Moralische Motivation: Kants Ethik in der Diskussion, Marburg, March 2004*, pp. 1-15
- Ryle, Gilbert (1946) 'Knowing How and Knowing That' *Proceedings of the Aristotelian Society* 46 pp. 1-16
- Ryle, Gilbert (1949) *The Concept of Mind* London: Hutchinson
- Ryle, Gilbert (1968) 'The Thinking of Thoughts: What is 'Le Penseur' Doing?' *University Lectures, The University of Saskatchewan* reprinted in Gilbert Ryle, 1971, *Collected Papers*, volume 2, London: Hutchinson, pp. 480-96
- Sayre-McCord, Geoff (2008) 'Metaethics', *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2008/entries/metaethics/>>.
- Scanlon, Tim (1998) *What We Owe to Each Other* Cambridge, MA: Belknap Press
- Schueler, George F. (1988) 'Modus Ponens and Moral Realism' *Ethics* 98 pp. 492-500
- Schueler, George F. (1995) *Desire: Its Role in Practical Reason and the Explanation of Action*, Cambridge, MA: MIT Press
- Schueler, George F. (2003) *Reasons and Purposes* Oxford: Oxford University Press



- Schneewind, Jerome B. (1998) 'Korsgaard and the Unconditional in Morality', *Ethics*, 109 (1) pp. 36-48
- Searle, John (1965) 'Meaning and Speech Acts' *Philosophical Review* 71 pp. 423-432
- Searle, John (1969) *Speech Acts* Cambridge: Cambridge University Press
- Shafer-Landau, Russ (1998) 'Moral Judgement and Moral Motivation' *The Philosophical Quarterly*, 48 (192) pp. 353-8
- Shafer-Landau, Russ (2003) *Moral Realism: A Defence* Oxford: Clarendon Press.
- Shafer-Landau, Russ (2006) 'Ethics as Philosophy: A Defence of Ethical Nonnaturalism' in Terry Horgan and Mark Timmons (eds.) 2006 *Metaethics after Moore* Oxford: Oxford University Press, pp. 209-32
- Shafer-Landau, Russ (2008) (ed.) *Oxford Studies in Metaethics, Volume 3* Oxford: Oxford University Press
- Singer, Peter (ed.) (1991) *A Companion to Ethics* Oxford: Blackwell
- Sinnott-Armstrong, Walter (1993) 'Some Problems for Gibbard's Norm-Expressivism' *Philosophical Studies* 69 pp. 297-313
- Sinnott-Armstrong, Walter and Timmons, Mark (eds.) (1996) *Moral Knowledge?* Oxford: Oxford University Press
- Skinner, B. F. (1953) *Science and Human Behavior* New York: MacMillan
- Sprigge, T. L. S. (1988) *The Rational Foundations of Ethics* London: Routledge
- Smit, Houston (2003) 'Internalism and the Origin of Rational Motivation' *The Journal of Ethics*, 7 pp. 183-231
- Smith, Michael (1987) 'The Humean Theory of Motivation', *Mind* 96 (381) pp. 36-61
- Smith, Michael (1988a) 'On Humeans, Anti-Humeans, and Motivation: A Reply to Pettit' *Mind*, 97 (388), pp. 589-595.
- Smith, Michael (1988b) 'Reason and Desire' *Proceedings of the Aristotelian Society*, 88 pp. 243-258
- Smith, Michael (1994) *The Moral Problem*, Oxford: Blackwell
- Smith, Michael (1995) 'Internal Reasons' *Philosophy and Phenomenological Research* 55 pp. 109-31
- Smith, Michael (1996a) 'The Argument for Internalism: reply to Miller', *Analysis* 56 pp. 175-83
- Smith, Michael (1996b) 'Normative Reasons and Full Rationality: Reply to Swanton', *Analysis* 56 pp. 160-68
- Smith, Michael (1997) 'In Defense of "The Moral Problem": A Reply to Brink, Copp, and Sayre-McCord', *Ethics*, 108 (1) pp. 84-119
- Smith, Michael (2003) 'Humeanism, Psychologism, and the Normative Story', *Philosophy and Phenomenological Research* LXVII (2) pp. 460-7
- Smith, Michael (2004a) *Ethics and the A Priori: Selected Essays on Moral Psychology and Meta-Ethics* Cambridge: Cambridge University Press
- Smith, Michael (2004b) 'Humean Rationality' in Alfred R. Mele and Piers Rawling (eds.) *The Oxford Handbook of Rationality*, Oxford: Oxford University Press, pp. 75-92
- Smith, Michael (2009) 'The Explanatory Role of Being Rational' in David Sobel and Steven Wall (eds.), *Reasons for Action*, New York: Cambridge University Press, 2009, pp.58-80
- Snare, Francis (1991) *Morals, Motivation, and Convention: Hume's Influential Doctrines* Cambridge: Cambridge University Press.
- Stanley, Jason and Williamson, Timothy (2001) 'Knowing How' *The Journal of Philosophy* 98 (8) pp. 411-44.

- Stevenson, Charles Leslie (1937) 'The Emotive Meaning of Ethical Terms' *Mind* 46 pp. 14-31
- Stevenson, Charles Leslie (1944) *Ethics and Language* New Haven and London: Yale University Press
- Stevenson, Charles Leslie (1963) *Facts and Values: Studies in Ethical Analysis* New Haven: Yale University Press
- Stich, Stephen (1983) *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press
- Stich, Stephen and Ravenscroft, Ian (1994) 'What is Folk Psychology?' *Cognition* 50 pp. 447-468
- Stocker, Michael (1979) 'Desiring the Bad: An Essay in Moral Psychology' *Journal of Philosophy* pp. 738-53
- Stoljar, Daniel (1993) 'Emotivism and Truth Conditions' *Philosophical Studies* 70 pp. 81-101
- Stratton-Lake, Philip and Hooker, Brad (2006) 'Scanlon versus Moore on Goodness' in Terry Horgan and Mark Timmons (eds.) 2006 *Metaethics After Moore* Oxford: Oxford University Press, pp. 149-68
- Stroud, Barry (1977) *Hume* London: Routledge and Kegan Paul
- Stroud, Sarah and Christine Tappolet (eds.) (2003) *Weakness of Will and Practical Irrationality* Oxford: Oxford University Press
- Superson, Anita M. (2009) *The Moral Skeptic* Oxford: Oxford University Press
- Svavarstóddir, Sigrún (1999) 'Moral Cognitivism and Motivation' *The Philosophical Review*, 108(2) pp. 161-219
- Svavarstóddir, Sigrún (2006) 'How Do Moral Judgements Motivate?' in James Dreier (ed.) *Contemporary Debates in Moral Theory* Oxford: Blackwell
- Swanton, Christine (1996) 'Is the Moral Problem Solved?' *Analysis* 56 pp. 155-60
- Talisse, Robert B. (2007) 'Does Value Pluralism Entail Liberalism?' *Social Science Research Network*, available at SSRN: <http://ssrn.com/abstract=1005651>, accessed 05/11/09
- Taylor, Charles (1964) *The Explanation of Behaviour* London: Routledge and Kegan Paul
- Thomson, Judith Jarvis (1997) 'The Right and the Good' *Journal of Philosophy*, 94 pp. 273-98
- Thomson, Judith Jarvis (2006) 'The Legacy of Principia' in Terry Horgan and Mark Timmons (eds.) 2006 *Metaethics after Moore* Oxford: Oxford University Press, pp. 233-54
- Unwin, Nicholas (1999) 'Quasi-Realism, Negation, and the Frege-Geach Problem' *Philosophical Quarterly* 49 pp. 337-52
- van Roojen, Mark (1996) 'Expressivism and Irrationality' *Philosophical Review* 105 pp. 311-35
- van Roojen, Mark (2000) 'Motivational Internalism: A Somewhat Less Idealized Account' *The Philosophical Quarterly* 50 pp. 233-241
- van Roojen, Mark (2002) 'Humean and Anti-Humean Internalism about Moral Judgements' *Philosophy and Phenomenological Research* 65 pp. 26-49
- van Roojen, Mark (2005) 'Expressivism, Supervenience and Logic' *Ratio* 17 pp. 190-205
- Velleman, David (1992) 'The Guise of the Good,' *Noûs* 26, pp. 3-26, reprinted in his *The Possibility of Practical Reason* Oxford: Oxford University Press, 2000
- Velleman, David (1996) 'The Possibility of Practical Reason', *Ethics* 106 pp. 694-726, reprinted in David Velleman (2000) *The Possibility of Practical Reason*.

- Velleman, David (2000) *The Possibility of Practical Reason*, Oxford: Oxford University Press
- Vinden, Penelope G. (1996) 'Junin Quechua Children's Understanding of Mind' *Child Development*, 67 pp.1707-16
- Vinden, Penelope G. (1999) 'Children's Understanding of Mind and Emotion: A Multi-culture Study' *Cognition & Emotion*, 13, pp. 19-48
- von Wright, Georg Henrik (1963) *The Varieties of Goodness* London: Routledge & Kegan Paul
- Wallace, R. Jay (1990) 'How to Argue about Practical Reason' *Mind*, 99 (395) pp. 355-385
- Wallace, R. Jay (1994) *Responsibility and the Moral Sentiments* Cambridge, MA: Harvard University Press
- Wallace, R. Jay (1999) 'Three Conceptions of Rational Agency' *Ethical Theory and Moral Practice*, 2 (3) pp. 217-242
- Wallace, R. Jay (2003) 'Explanation, Deliberation, and Reasons' *Philosophy and Phenomenological Research*, LXVII (2) pp. 429-435
- Wallace, R. Jay; Pettit, Philip; Scheffler, S. and Smith, Michael (eds.) (2004) *Reason and Value: Themes from the Moral Philosophy of Joseph Raz* Oxford: Clarendon Press
- Wallace, R. Jay (2006a) *Normativity and the Will: Selected Essays on Moral Psychology and Practical Reason* Oxford: Oxford University Press
- Wallace, R. Jay (2006b) 'Moral Motivation' in James Dreier (ed.) (2006) *Contemporary Debates in Moral Theory* Oxford: Blackwell
- Watson, Gary (1975) 'Free Agency' reprinted in Gary Watson (ed.) 1982 *Free Will* Oxford: Oxford University Press, pp. 96-110
- Watson, Gary (1982) *Free Will* Oxford: Oxford University Press
- Watson, Gary (2003) 'The Work of the Will', in Stroud and Tappolet (eds.) *Weakness of Will and Practical Irrationality* Oxford: Oxford University Press, pp. 172-200
- Wedgwood, Ralph (1997) 'Non-cognitivism, Truth and Logic' *Philosophical Studies* 86 pp. 73-91
- Wedgwood, Ralph (2002) 'Practical Reason and Desire' *Australasian Journal of Philosophy* 80 pp. 345-358
- Widlock, Thomas (2004) 'Sharing by Default? Outline of an Anthropology of Virtue', *Anthropological Theory* 4 (1) pp. 53-70
- Wiggins, David (1987) 'A Sensible Subjectivism?' in his *Needs, Values, Truth* Oxford: Basil Blackwell, pp. 185-214
- Wiggins, David (1990-1991) 'Moral Cognitivism, Moral Relativism and Motivating Moral Beliefs' *Proceedings of the Aristotelian Society*, New Series, 91 pp. 61-85
- Wiggins, David (2006) *Ethics: Twelve Lectures on the Philosophy of Morality* London: Penguin
- Williams, Bernard (1976) 'Persons, Character and Morality' reprinted in Bernard Williams 1981 *Moral Luck*, Cambridge: Cambridge University Press, pp. 1-19
- Williams, Bernard (1979) 'Internal and External Reasons' in Ross Harrison (ed.) *Rational Action: Studies in Philosophy and Social Science* Cambridge: Cambridge University Press.
- Williams, Bernard (1981) *Moral Luck*, Cambridge: Cambridge University Press
- Williams, Bernard (1985) *Ethics and the Limits of Philosophy*, Hammersmith, London: Fontana Press

- Williams, Bernard (1995) *Making Sense of Humanity* Cambridge: Cambridge University Press
- Williams, Bernard (1996) 'History, Morality and the Test of Reflection' in Onora O'Neill (ed.) *The Sources of Normativity* Cambridge: Cambridge University Press
- Williamson, J. (1970) 'Pro- and Con- Attitudes' *The Philosophical Quarterly*, 20, pp. 357-367.
- Wilson, George (1989) *The Intentionality of Human Action* Stanford: Stanford University Press
- Wittgenstein, Ludwig (1972) *Philosophical Investigations*, trans. G. E. M. Anscombe, Oxford: Blackwell
- Wodehouse, Helen (1938) 'Language and Moral Philosophy' *Mind* 47 pp. 200-13
- Wolf, Susan (2009) 'One Thought Too Many: Love, Morality, and the Ordering of Commitment', presented at *Themes from the Ethics of Bernard Williams*, University of Leeds, 30<sup>th</sup> June – 2<sup>nd</sup> July 2009
- Woods, Michael (1972) 'Reasons for Action and Desires' *Proceedings of the Aristotelian Society, Supplementary Volumes*, 46 pp. 189-201
- Wright, Crispin (1988) 'Realism, Antirealism, Irrealism, Quasi-Realism' *Midwest Studies in Philosophy* 12 pp. 25-49
- Wright, Crispin (1992) *Truth and Objectivity* Cambridge, MA: Harvard University Press
- Wright, Larry (1976) *Teleological Explanations: An Etiological Analysis of Goals and Functions* Berkley: University of California Press
- Yablo, Stephen (2001) 'Coulda, Woulda, Shoulda' in Tamar Szabó Gendler and John Hawthorne (eds.) (2001) *Conceivability and Possibility* Oxford: Oxford University Press
- Zangwill, Nick (1998) 'Direction of Fit and Normative Functionalism' *Philosophical Studies* 91 pp. 173–203
- Zangwill, Nick (2003) 'Externalist Moral Motivation' *American Philosophical Quarterly* 40 (2) pp. 143-54
- Zangwill, Nick (2008a) 'Besires and the Motivation Debate' *Theoria* 74 pp. 50-9
- Zangwill, Nick (2008b) 'The Indifference Argument' *Philosophical Studies* 138 pp. 91-124
- Zangwill, Nick (2008c) 'Moral Dependence' in Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics, Volume 3*, Oxford: Oxford University Press, pp.109-127
- Ziff, Paul (1960) *Semantic Analysis* Ithaca, NY: Cornell University Press
- Zimmerman, Michael J. (2001) *The Nature of Intrinsic Value* Oxford: Rowman & Littlefield